

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: Brenwen Loeb Examiner #: 78225 Date: 2/2/01
 Art Unit: 1636 Phone Number 301 605-1197 Serial Number: 09/000,002
 Mail Box and Bldg/Room Location: 411A-09 Results Format Preferred (circle): PAPER DISK E-MAIL
11E-12

If more than one search is submitted, please prioritize searches in order of need.

 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Nucleic acid fragments recombinant vector containing the same and method for
promoting the expression of structural gene by using the same
 Inventors (please provide full names): _____

→ UEKI, Jun

Earliest Priority Filing Date: 11/19/98

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search SEQ. ID. NOS. 1-3 in the
 nucleotide databases. Don't search pending
 patents. Thank you.

POINT OF CONTACT:
 BARB O'BRYEN
 TECH. INFORMATION SPECIALIST
 STIC CM1 12C14 308-4291

STAFF USE ONLY

Searcher: BOB
 Searcher Phone #: _____
 Searcher Location: _____
 Date Searcher Picked Up: 2-5-01
 Date Completed: 2-13-01
 Searcher Prep & Review Time: 15
 Clerical Prep Time: _____
 Online Time: 12

Type of Search

NA Sequence (#) 3
 AA Sequence (#) _____
 Structure (#) _____
 Bibliographic _____
 Litigation _____
 Fulltext _____
 Patent Family _____
 Other _____

Vendors and cost where applicable

STN _____
 Dialog _____
 Questal/Orbit _____
 Dr.Link _____
 Lexis/Nexis _____
 Sequence Systems _____
 WWW/Internet _____
 Other (specify) _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2001, 20:28:44 : Search time 136.04 Seconds
(without alignments)
176,730 Million cell updates/sec

Title: US-09-600-602-1

Perfect score: 64
1 taagcccaagtgtcttaagc.....gcttctcttcgcgcaca 64

Sequence: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Scoring table: 480022 segs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_36:*

1: /SID56/gcgdata/geneseq/NA1980.DAT:*

2: /SID56/gcgdata/geneseq/NA1981.DAT:*

3: /SID56/gcgdata/geneseq/NA1982.DAT:*

4: /SID56/gcgdata/geneseq/NA1983.DAT:*

5: /SID56/gcgdata/geneseq/NA1984.DAT:*

6: /SID56/gcgdata/geneseq/NA1985.DAT:*

7: /SID56/gcgdata/geneseq/NA1986.DAT:*

8: /SID56/gcgdata/geneseq/NA1987.DAT:*

9: /SID56/gcgdata/geneseq/NA1988.DAT:*

10: /SID56/gcgdata/geneseq/NA1989.DAT:*

11: /SID56/gcgdata/geneseq/NA1990.DAT:*

12: /SID56/gcgdata/geneseq/NA1991.DAT:*

13: /SID56/gcgdata/geneseq/NA1992.DAT:*

14: /SID56/gcgdata/geneseq/NA1993.DAT:*

15: /SID56/gcgdata/geneseq/NA1994.DAT:*

16: /SID56/gcgdata/geneseq/NA1995.DAT:*

17: /SID56/gcgdata/geneseq/NA1996.DAT:*

18: /SID56/gcgdata/geneseq/NA1997.DAT:*

19: /SID56/gcgdata/geneseq/NA1998.DAT:*

20: /SID56/gcgdata/geneseq/NA1999.DAT:*

21: /SID56/gcgdata/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	100.0	173	17	Promoter #2 for ph
2	64	100.0	173	19	ubiquitin gene int
3	64	100.0	183	17	Promoter #1 for ph
4	64	100.0	2799	16	DNA encoding Phosp
5	64	100.0	2799	17	Phospholipase D ge
6	24.8	38.8	1331	8	DNA sequence encod
7	24.8	38.8	2110	19	Mouse Friend virus
8	23.6	36.9	1037	20	Human secreted pro
9	23.6	36.9	2306	11	Sequence encoding
10	23.6	36.9	2306	11	Murine IL-2R beta
11	23.6	36.9	7947	20	Enterococcus faeca
12	23.4	36.6	508	18	Hamster Ubiquitin/

13	22.8	35.6	219	20	Z20473
14	22.8	35.6	2353	19	V455981
15	22.8	35.6	2464	18	V03752
16	22.8	35.6	5888	20	Z24013
17	22.6	35.3	3155	17	T42120
18	22.6	35.3	3155	21	Z51507
19	22.4	35.0	3336	6	MS0149
20	22.2	34.7	332	16	T19131
21	22.2	34.7	1738	19	V09299
22	22.2	34.7	9558	16	O88328
23	21.8	34.1	380	20	V86878
24	21.8	34.1	1228	20	X82099
25	21.8	34.1	1608	21	Z22298
26	21.8	34.1	2417	20	X91430
27	21.8	34.1	2417	20	X91431
28	21.6	33.8	2059	19	V52041
29	21.4	33.4	1551	21	Z98176
30	21.4	33.4	2097	19	Z96306
31	21.4	33.4	2224	20	Z10617
32	21.4	33.4	2948	20	X18940
33	21.4	33.4	3071	19	V65214
34	21.4	33.4	4060	20	X20655
35	21.4	33.4	9927	20	X12997
36	21.4	33.4	20199	19	V52139
37	21.4	33.4	49999	20	Z23899
38	21.2	33.1	873	19	V47586
39	21.2	33.1	2277	19	V13836
40	21.2	33.1	2277	19	V05372
41	21.2	33.1	7198	20	X60278
42	21.2	33.1	99960	21	Z50905
43	21	32.8	340	20	X41055
44	21	32.8	598	21	Z80454
45	21	32.8	607	21	Z80335

ALIGNMENTS

RESULT 1

T42852 standard; DNA: 173 BP.

AC T42852:

16-JUN-1997 (first entry)

DE Promoter #2 for phospholipase D gene.

DE XX

KW Phospholipase D; rice; promoter; ss.

KW XX

OS Oryza sativa.

OS XX

PN W09630510-A1.

PN XX

PD 03-OCT-1996.

PD XX

PE 28-MAR-1996; 96MO-JP00812.

PE XX

PR 29-MAR-1995; 95JP-0096126.

PR XX

PA (N1SB) JAPAN TOBACCO INC.

PA XX

PI Moritoka S, Ueki J:

PI XX

DR WPI; 1996-455357/45.

DR XX

PT Promoter DNA sequence derived from rice - used to increase expression of foreign genes in transformed hosts

PT XX

PS Claim 3; Page 23; 29pp; Japanese.

PS XX

CC T42851 and T42852 represent promoters isolated from rice. These sequences are specifically promoters for the phospholipase D gene (PLD). This

Barley microsatell

A. thaliana sterol

Porcine TNFalpha-c

Human GDNF fragmen

M-lats2 gene encod

Mouse lats2 (large

Sequence of the cd

Human gene signatu

Nucleotide sequenc

Valencia orange rl

EST clone BK241.

Human calcium-acti

Human potassium ch

T. gondii MG154-8

T. gondii MG154-8

Helicobacter poly

Human signal pepti

S. pneumoniae derl

cdNA encoding a mu

Human basic helix-

DNA encoding a S.

Polynucleotide seq

Enterococcus faeca

Streptococcus pneu

Human LOBO homolog

Leishmania antigen

Homo sapiens mamma

Human telomerase p

DNA encoding a 2,4

Human TBC-1 partia

Human secreted pro

Human colon cancer

Human colon cancer

CC sequence represents a shortened version of the promoter sequence shown in
CC T42851. These sequences are efficient promoters for greatly increasing
CC the expression of foreign genes in transformant rice and other plants.
XX
SQ Sequence 173 BP; 21 A; 48 C; 47 G; 57 T; 0 other;

Query Match 100.0%; Score 64; DB 17; Length 173;
Best Local Similarity 100.0%; Pred. No. 9.1e-15;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taagccagtgcttaggctaagcgacactagagctctctgctgctctctccg 60
|||||
DB 2 taagccagtgcttaggctaagcgacactagagctctctgctgctctctccg 61

OY 61 ctca 64
||||
DB 62 ctca 65

RESULT 2
V05554
ID V05554 standard; DNA; 173 BP.

AC V05554;
DT 18-JUN-1998 (first entry)
DE Ubiquitin gene Intron.

XX
XX
KM Intron; Oryza sativa; ubiquitin gene; ds.
XX

OS Oryza sativa.
XX
PN WO9747755-A1.

PD 18-DEC-1997.
XX
PF 12-JUN-1997; 97WO-JP02030.

XX
PR 12-JUN-1996; 96JP-0172922.
XX
PA (NISR) JAPAN TOBACCO INC.

XX
PI Kureya Y, Morioka S, Ohta S, Ueki J;
XX
DR WPI; 1998-086573/08.

PT Method for expressing foreign gene - comprises inserting gene
downstream from promoter comprising at least two introns, useful in,
e.g. genetic engineering
XX
PS Claim 4; Page 10; 20pp; Japanese.

CC This sequence represents an Intron from the Oryza sativa ubiquitin gene,
CC and can be used in the method of the invention. The method is for
CC expressing a foreign gene by inserting the gene downstream from a
CC promoter, comprising at least 2 Intron sequences. The method is used for
CC expression of foreign genes, useful in, e.g. genetic engineering. The
CC method achieves higher levels of expression than conventional methods.
XX
SQ Sequence 173 BP; 21 A; 48 C; 47 G; 57 T; 0 other;

Query Match 100.0%; Score 64; DB 19; Length 173;
Best Local Similarity 100.0%; Pred. No. 9.1e-15;

Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taagccagtgcttaggctaagcgacactagagctctctgctgctctctccg 60
|||||
DB 2 taagccagtgcttaggctaagcgacactagagctctctgctgctctctccg 61

OY 61 ctca 64

DB 62 ctca 65

||||

RESULT 3
T42851
ID T42851 standard; DNA; 183 BP.

XX
AC T42851;
XX
DT 16-JUN-1997 (first entry)
DE Promoter #1 for phospholipase D gene.

XX
KM Phospholipase D; rice; promoter; ss.
XX
OS Oryza sativa.

XX
PN WO9630510-A1.
PD 03-OCT-1996.

XX
PF 28-MAR-1996; 96WO-JP00812.
PR 29-MAR-1995; 95JP-0096126.

XX
PA (NISR) JAPAN TOBACCO INC.
XX
PI Morioka S, Ueki J;
XX
DR WPI; 1996-455357/45.

PT Promoter DNA sequence derived from rice - used to increase
expression of foreign genes in transformed hosts
XX
PS Claim 1; Page 14; 29pp; Japanese.

XX
CC T42851 and T42852 represent promoters isolated from rice. These sequences
CC are specifically promoters for the phospholipase D gene (PLD). T42852
CC represents a shortened version of this sequence. These sequences are
CC efficient promoters for greatly increasing the expression of foreign
CC genes in transformant rice and other plants.
XX
SQ Sequence 183 BP; 23 A; 51 C; 51 G; 58 T; 0 other;

Query Match 100.0%; Score 64; DB 17; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.3e-15;

Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taagccagtgcttaggctaagcgacactagagctctctgctgctctctccg 60
|||||
DB 7 taagccagtgcttaggctaagcgacactagagctctctgctgctctctccg 66

OY 61 ctca 64
||||
DB 67 ctca 70

RESULT 4
Q86785
ID Q86785 standard; DNA; 2799 BP.

XX
AC Q86785;

XX
DT 04-MAR-1996 (first entry)
DE DNA encoding Phospholipase D.

XX
KM phospholipase D; measurement; reagent; phospholipid level; ss.

XX
OS Zea mays.


```

FH Key Location/Qualifiers
FT exon 1876..1983
FT /*tag= a
FT intron 1984..2523
FT /*tag= b
FT exon 2524..2799
FT /*tag= c
FT /note= "partial exon 2"

XX MO9509234-A1.
XX
XX PD 06-APR-1995.
XX
XX PF 30-SEP-1994; 94WO-JP01627.
XX
XX PR 30-SEP-1993; 93JP-0267884.
XX
XX PA (NISB ) JAPAN TOBACCO INC.
XX
XX PI Morioka S, Ueki J;
XX
XX DR WPI: 1995-147433/19.
XX
XX DR P-PSDB: R72799.
XX
XX PT Cloned DNA coding plant derived phospholipase D - controls
XX
XX PT expression of plant derived PLD gene
XX
XX PS Claim 14; Page 33-35; 41pp: Japanese.
XX
XX CC The DNA contains sequences necessary for the expression of a plant
CC derived phospholipase D (PLD). The PLD is useful for measuring
CC phospholipid levels and for producing derivs. by e.g. base exchange
CC reactions.
XX
XX SQ Sequence 2799 BP; 692 A; 709 C; 609 G; 789 T; 0 other;

Query Match 100.0%; Score 64; DB 16; Length 2799;
Best Local Similarity 100.0%; Pred. No. 1.7e-14;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 taagccagtgcttaagcctaagcgacactagagcttctgcgtctcttcctccg 60
    |||||||
Db 1667 taagccagtgcttaagcctaagcgacactagagcttctgcgtctcttcctccg 1726

OY 61 ctca 64
    ||||
Db 1727 ctca 1730

RESULT 5
T42854
ID T42854 standard; cDNA to mRNA; 2799 BP.
XX
XX AC T42854;
XX
XX DT 16-JUN-1997 (first entry)
XX
XX DE Phospholipase D gene sequence fragment.
XX
XX KW Phospholipase D; rice; promoter; ss.
XX
XX OS Oryza sativa.
XX
XX FH key
XX FT exon 1876..1983
XX FT /*tag= a
XX FT /number= 1
XX FT /note= "encodes residues 1 to 36 of W06134"
XX FT intron 1984..2523
XX FT /*tag= b
XX FT /number= 1
XX FT exon 2524..2799
```

```

FT /*tag= c
FT /number= 2
FT /note= "encodes residues 37 to 128 of W06134"

XX MO9630510-A1.
XX
XX PD 03-OCT-1996.
XX
XX PF 28-MAR-1996; 96WO-JP00812.
XX
XX PR 29-MAR-1995; 95JP-0096126.
XX
XX PA (NISB ) JAPAN TOBACCO INC.
XX
XX PI Morioka S, Ueki J;
XX
XX DR WPI: 1996-455357/45.
XX
XX DR P-PSDB: W06134.
XX
XX PT Promoter DNA sequence derived from rice - used to increase
XX
XX PT expression of foreign genes in transformed hosts
XX
XX PS Disclosure; Page 20-22; 29pp: Japanese.
XX
XX CC This sequence represents a fragment of the coding sequence of the rice
XX phospholipase D gene (PLD). The promoter for the PLD gene was isolated
XX using the primers shown in T42857 and T42588. The promoters (see T42851
XX and T42852) are efficient promoters for greatly increasing the expression
XX of foreign genes in transformant rice and other plants.
XX
XX SQ Sequence 2799 BP; 692 A; 709 C; 609 G; 789 T; 0 other;

Query Match 100.0%; Score 64; DB 17; Length 2799;
Best Local Similarity 100.0%; Pred. No. 1.7e-14;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 taagccagtgcttaagcctaagcgacactagagcttctgcgtctcttcctccg 60
    |||||||
Db 1667 taagccagtgcttaagcctaagcgacactagagcttctgcgtctcttcctccg 1726

OY 61 ctca 64
    ||||
Db 1727 ctca 1730

RESULT 6
N70891
ID N70891 standard; DNA; 1331 BP.
XX
XX AC N70891;
XX
XX DT 26-FEB-1991 (first entry)
XX
XX DE DNA sequence encoding tumor necrosis factor.
XX
XX KW Tumor necrosis factor; ss; antitumor.
XX
XX FH key
XX FT mat_peptide 424..885
XX FT /*tag= a
XX FT /label= tumor necrosis factor protein

XX PN JP62135493-A.
XX
XX PD 18-JUN-1987.
XX
XX PF 06-DEC-1985; 85JP-0275392.
XX
XX PR 06-DEC-1985; 85JP-0275392.
XX
XX PA (TANA ) TANABE SEIYAKU KK.
```


CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
CC disorders, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
CC The sequences given in X30174 to X30182 and Y04322 to Y04334 are used
CC in the exemplification of the present invention.

XX Sequence 1037 BP; 308 A; 198 C; 224 G; 306 T; 1 other;

Query Match 36.9%; Score 23.6; DB 20; Length 1037;

Best Local Similarity 69.6%; Pred. No. 8.6;

Matches 32; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Y 18 ggcctagcgcactagctcttgcctgcctgcctcttcgcgcgc 63
||| | |||| | | | |||| | |||| | ||| | |||
Db 14 ggcctgtggcactggaactagctgctgctgctacatcaagtc 59

RESULT 9

ID Q05873 standard; DNA; 2306 BP.

XX Q05873;

AC Q05873;

XX Q05873;

DT 07-JAN-1991 (first entry)

XX 07-JAN-1991 (first entry)

DE Sequence encoding murine interleukin-2 receptor beta chain.

XX Sequence encoding murine interleukin-2 receptor beta chain.

XX IL-2R beta: pIL-2R beta 9; ss;

KW IL-2R beta: pIL-2R beta 9; ss;

XX IL-2R beta: pIL-2R beta 9; ss;

OS Mus musculus.

XX Mus musculus.

XX Mus musculus.

FT Key Location/Qualifiers

FT CDS 375..1994

FT /*tag= a

XX EP386304-A.

XX EP386304-A.

PD 12-SEP-1990.

XX 12-SEP-1990.

XX 29-MAY-1989; 89EP-0109656.

XX 29-MAY-1989; 89EP-0109656.

PR 29-MAY-1989; 89EP-0109656.

PR 07-MAR-1989; 89EP-0104023.

XX 07-MAR-1989; 89EP-0104023.

XX (OSAU) OSAKA UNIVERSITY.

XX (OSAU) OSAKA UNIVERSITY.

PI Taniguchi T;

XX Taniguchi T;

DR WPI: 1990-276456/37.

XX WPI: 1990-276456/37.

DR P-PSDB: R06647.

XX P-PSDB: R06647.

XX Recombinant interleukin-2 receptor beta chain - used for studying

XX Recombinant interleukin-2 receptor beta chain - used for studying

XX IL-2 system and producing antibodies for diagnosis and therapy.

XX IL-2 system and producing antibodies for diagnosis and therapy.

XX Claim 6; Fig 8; 37pp; English.

XX Claim 6; Fig 8; 37pp; English.

XX IL-2 receptor beta chains are useful in studying the biochemistry of

XX IL-2 receptor beta chains are useful in studying the biochemistry of

CC interleukin, and in diagnosis and therapy by immune suppression and

CC interleukin, and in diagnosis and therapy by immune suppression and

XX activation.

XX activation.

XX Sequence 2306 BP; 482 A; 688 C; 585 G; 551 T; 0 other;

XX Sequence 2306 BP; 482 A; 688 C; 585 G; 551 T; 0 other;

Query Match 36.9%; Score 23.6; DB 11; Length 2306;

Best Local Similarity 76.3%; Pred. No. 10;

Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

RESULT 10
ID Q06332 standard; cDNA; 2306 BP.
XX Q06332
AC Q06332;

XX Q06332;

XX Q06332;

DT 07-FEB-1991 (first entry)

XX 07-FEB-1991 (first entry)

XX Murine IL-2R beta chain.

DE Murine IL-2R beta chain.

XX Interleukin; receptor; ss.

KW Interleukin; receptor; ss.

XX Interleukin; receptor; ss.

OS Mus musculus.

XX Mus musculus.

XX Mus musculus.

FT Key Location/Qualifiers

FT CDS 375..1994

FT /*tag= a

FT misc_difference 1182..1184

FT /*tag= b

FT /*tag= b

XX /*tag= b

XX /*tag= b

PN AU950726-A.

XX AU950726-A.

XX 13-SEP-1990.

PD 13-SEP-1990.

XX 13-SEP-1990.

XX 06-MAR-1990; 90AU-0050726.

PR 06-MAR-1990; 90AU-0050726.

PR 20-JUL-1989; 89EP-0113310.

PR 07-MAR-1989; 89EP-0104023.

PR 29-MAY-1989; 89EP-0109656.

XX 29-MAY-1989; 89EP-0109656.

XX (BOEH) BOEHRINGER INGELHEI.

XX (BOEH) BOEHRINGER INGELHEI.

PI Taniguchi T, Hatakeyama M, Minamoto S, Kono T, Doi T;

XX Taniguchi T, Hatakeyama M, Minamoto S, Kono T, Doi T;

PI Miyasaka M, Tsudo, Karasuyama H;

XX Miyasaka M, Tsudo, Karasuyama H;

DR WPI: 1990-327673/44.

XX WPI: 1990-327673/44.

DR P-PSDB: R07507.

XX P-PSDB: R07507.

XX Recombinant interleukin-2 receptor beta chain - useful for

XX Recombinant interleukin-2 receptor beta chain - useful for

XX diagnosis and therapy by immune suppression or activation.

XX diagnosis and therapy by immune suppression or activation.

XX Claim 4; Fig 8; 65pp; English.

XX Claim 4; Fig 8; 65pp; English.

XX The sequence was obtd. from clones isolated from a cDNA library

XX The sequence was obtd. from clones isolated from a cDNA library

CC prepd. from RNA extracted from Concanavalin A-stimulated mouse

CC prepd. from RNA extracted from Concanavalin A-stimulated mouse

CC spleen cells. The clones were selected using a probe derived from

CC spleen cells. The clones were selected using a probe derived from

CC the human sequence. The sequence encodes a soluble portion of the

CC the human sequence. The sequence encodes a soluble portion of the

CC IL-2R beta chain.

XX IL-2R beta chain.

XX See also Q06331.

XX See also Q06331.

XX Sequence 2306 BP; 487 A; 685 C; 580 G; 550 T; 4 other;

XX Sequence 2306 BP; 487 A; 685 C; 580 G; 550 T; 4 other;

Query Match 36.9%; Score 23.6; DB 11; Length 2306;

Best Local Similarity 76.3%; Pred. No. 10;

Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Y 25 cgcactagctcttgcctgcctgcctcttcgcgcgc 62
| | | | | | | | | | | | | | | | | | | | |
Db 9 ctctcttgccttgcctgcctgcctgcctcttgcgcgc 46

RESULT 11

X13021/c

ID X13021 standard; DNA; 7947 BP.

XX X13021;

AC X13021;

XX X13021;

DT 19-MAR-1999 (first entry)

XX 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome config SEQ ID NO:84.

```

XX Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX Enterococcus faecalis.
XX WO9850555-A2.
XX 12-NOV-1998.
XX 04-MAY-1998; 98MO-US08985.
XX 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Barash SC, Dillon PJ, Kunsch CA;
XX WPI; 1999-045171/04.
XX
XX New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
XX Claim 1; Page 580-584; 2084pp; English.
XX
XX A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
XX Sequence 7947 BP; 2720 A; 1379 C; 1741 G; 2100 T; 7 other:
SQ

```

```

Query Match          36.9%; Score 23.6; DB 20; Length 7947;
Best Local Similarity 69.6%; Pred. NO. 14;
Matches 32; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY      8 agtgcctagagcctaagcagcactagagctctctgcgctgcttc 53
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      4084 AGTTAGCTTCGTAAACGTTGCTTCTTCTATGTCGCCCTTCTTC 4039

```

```

RESULT 12
T87466/C
ID T87466 standard; DNA; 508 BP.
XX
AC T87466;
XX
DT 29-JAN-1998 (first entry)
XX
DE Hamster Ubiquitin/S27a cDNA sequence.
XX
KW CHO; hamster; Ub/S27a; promoter; ubiquitin; primer; ss.
XX
OS Hamster.
XX
XX Key Location/Qualifiers
FH 17.487
FT CDS
FT /*tag- a

```

```

FT misc_RNA
FT 1..244
FT /*tag- b
FT /note- "ubiquitin"
FT Intron
FT 64..65
FT /*tag- c
FT /note- "Site of intron"
FT 204..205
FT /*tag- d
FT /note- "Site of intron"
FT Intron
FT 337..338
FT /*tag- e
FT /note- "Site of intron"
FT polyA_signal
FT 497..502
FT /*tag- f
FT polyA_signal
FT 501..506
FT /*tag- g
XX
PN DE19539493-A1.
XX
PD 30-APR-1997.
XX
PF 24-OCT-1995; 95DE-1039493.
XX
PR 24-OCT-1995; 95DE-1039493.
XX
PA (THOM ) THOMAE GMBH KARL.
XX
PI Bergemann K, Enekel B, Gannon F, Noe W;
XX
DR WPI; 1997-246232/23.
DR P-PSDB; W28542.
XX
XX Nucleic acid containing promoter and regulatory regions of the
PT hamster ubiquitin S27a gene - useful for production of heterologous
PT proteins, especially in CHO cells
XX
PS Claim 30; Fig 1; 22pp; German.
XX
XX The cDNA sequences of Ub/S27a isolated from CHO cells (T87466)
CC and from human cells (T87467) were compared. They show 92.2%
CC homology. The corresponding amino acid sequences show 100%
CC homology.
XX
XX Sequence 508 BP; 159 A; 98 C; 130 G; 121 T; 0 other:
SQ

```

```

Query Match          36.6%; Score 23.4; DB 18; Length 508;
Best Local Similarity 67.3%; Pred. NO. 8.6;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY      11 gtcctagagcctaagcagcactagagctctctgcgctgcttc 59
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      331 GTACTTAGCAGCAGCAGCACTTAACCTTCTTCTATGCTTATCTTC 283

```

```

RESULT 13
Z20473
ID Z20473 standard; DNA; 219 BP.
XX
AC Z20473;
XX
DT 19-NOV-1999 (first entry)
XX
DE Barley microsatellite marker clone Bmag5.
XX
KW Microsatellite marker; barley; chromosome 7 marker; cereal;
KW fermentability; group 5 chromosome; ethyl carbanate production; Bmac213;
KW wort fermentation; Triticaceae; Bmac96; epi-heterodendrin production;
XX diagnosis; ss.
XX
XX Hordeum vulgare.
OS
XX
PN WO946404-A1.

```

```

XX PD 16-SEP-1999.
XX PF 01-MAR-1999; 99WO-GB00602.
XX PR 10-MAR-1998; 98GB-0005087.
XX PA (SCCR-) SCOTTISH CROP RES INST.
XX PI Thomas WTB, Swanston JS, Powell W, Maugh R, Ramsey LD;
XX WPI: 1999-551424/46.
XX DR
XX PT Screening cereals for fermentability, especially useful in barley -
XX PS Disclosure: Fig 7; 49pp; English.
XX CC This sequence represents a barley microsatellite marker. Primers that
XX CC amplify this sequence can be used in the method of the invention.
XX CC The method is for screening cereal for fermentability, comprising
XX CC analysing cereal genomic DNA to determine which allele(s) of a gene/gene
XX CC complex affecting fermentability at a locus close to the centromere on
XX CC homologous Triticeae group 5 chromosome (barley chromosome 7) is/are
XX CC present. The invention also relates to a method for screening cereal for
XX CC ethyl carbamate production on wort fermentation and distillation.
XX CC comprising analysing barley genomic DNA to determine which allele(s) of
XX CC the locus, designated eph on the short arm of homologous Triticeae
XX CC group 1 chromosome (barley chromosome 5) is/are present. The methods and
XX CC primers are useful for identifying microsatellites Bmac96 and Bmac213,
XX CC which are useful for determining fermentability and/or epi-heterodendrin
XX CC production in cereals, especially barley. Current methods for determining
XX CC fermentability are difficult to apply within barley breeding programs.
XX CC Prior art methods using molecular markers have difficulty in detecting
XX CC levels of allelic variation.
XX SQ Sequence 219 BP; 42 A; 48 C; 45 G; 84 T; 0 other:

Query Match          35.6%; Score 22.8; DB 20; Length 219;
Best Local Similarity 66.0%; Pred. No. 12;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 9 gtgtgcttagcgaagcactagactctctgctgctgtctcttc 58
    ||||| ||| | | | | |||| | ||| | |||| |||||
DB 51 gtgtgcatagacctacaggtccatagctagctagctgctcttc 100

RESULT 14
V45981
ID V45981 standard; DNA: 2353 BP.
XX AC V45981;
XX DT 09-OCT-1998 (first entry)
XX DE A. thaliana sterol glycosyltransferase cDNA.
XX KW Sterol glycosyl transferase; oat; transgenic plant; tolerance;
XX KW resistance; environmental factor; drought; salt; cold; frost;
XX KW fungal attack; temperature; secondary metabolite; inhibitor; antitumour;
XX KW vascular permeability; anti-inflammatory; haemostatic agent; ss.
XX OS Arabidopsis thaliana.
XX FH
XX FT CDS Key Location/Qualifiers
    113..2026
    /tag= a
    /product= SGT
    /note= "sterol glycosyltransferase"
XX PN W09817789-A1.
XX PD 30-APR-1998.

```

```

XX PF 10-OCT-1997; 97WO-DE02335.
XX PR 21-OCT-1996; 96DE-1043309.
XX PA (BALT/) BALTRUSCH A.
XX PA (BALT/) BALTRUSCH R M.
XX PA (GVSE-) GVS GES ERWERB & PA VERWERTUNG LANDWIRTSCHA.
XX PI Baltrusch M, Heinz E, Warnecke D, Wolter FP;
XX WPI: 1998-271789/24.
XX DR P-PSDB: W64388, W64389
XX PT DNA encoding sterol glycosyltransferase and related enzymes -
XX PT transgenic plants, with increased resistance to stress, and
XX PT organisms, used to produce steryl glycoside(s), potentially useful
XX PT therapeutically, in high yield
XX PS Example 9; Fig 17; 64pp; German.
XX CC This sequence encodes a sterol glycosyltransferase (SGT) isolated from
XX CC A. thaliana which is used in a method to produce transgenic plants which
XX CC have a better tolerance of/resistance against adverse environmental
XX CC factors such as drought, high salt content of the soil, cold/frost and
XX CC fungal attack using recombinant SGT. Microorganisms containing a
XX CC recombinant SGT protein have better resistance to high salt/ethanol
XX CC content of growth media, cold/frost and high temperature. Organisms
XX CC containing such recombinant proteins are used to produce secondary
XX CC metabolites which are potentially useful for inhibiting vascular
XX CC permeability, or as antitumour, anti-inflammatory and haemostatic agents.
XX SQ Sequence 2353 BP; 700 A; 468 C; 518 G; 667 T; 0 other:

Query Match          35.6%; Score 22.8; DB 19; Length 2353;
Best Local Similarity 71.4%; Pred. No. 20;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 18 ggctagcgacagcagctctgtctgctgtctcttc 59
    ||||| ||| | | | ||||| |||| | ||||| |||
DB 130 ggctgagctgcgcacaggtctctctcgtctctctcttc 171

RESULT 15
V03752
ID V03752 standard; CDNA: 2464 BP.
XX AC V03752;
XX DT 15-APR-1998 (first entry)
XX DE Porcine TNFalpha-convertase coding sequence.
XX KW Tumour necrosis factor alpha convertase; TNFalpha-convertase; human;
XX KW inhibitor; therapy; systemic inflammatory response syndrome; malignancy;
XX KW reperfusion injury; cardiovascular disease; infectious disease; cachexia;
XX KW obstetrical disorder; gynaecological disorder; inflammatory disease; pig;
XX KW autoimmunity; allergic disease; atopic disease; septic shock; psoriasis;
XX KW graft osteoporosis; restenosis; rheumatoid arthritis; osteoarthritis;
XX KW macular degeneration; multiple sclerosis; AIDS; porcine; ss.
XX OS Sus scrofa.
XX FH
XX FT CDS Key Location/Qualifiers
    1..1311
    /tag= a
XX PN W09735538-A2.
XX PD 02-OCT-1997.
XX PF 25-MAR-1997; 97WO-EP01497.

```

```

XX 26-MAR-1996; 96US-0620663.
PR
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
PI Becherer JD, Chen W, Didsbury JR, Jin SC, McGeehan GM;
PI Moss ML, Rocque WJ, Schoenen FJ;
XX
DR WPI: 1997-489331/45.
P-PSDB: W40996.
XX
PT Mammalian tumour necrosis factor alpha convertase - useful to screen
PT for new inhibitors which can treat disorders involving abnormal
PT levels of TNF alpha, e.g. inflammatory and cardiovascular disease
XX
PS Example: Fig 6; 132pp; English.
XX
CC This sequence represents the coding sequence for the porcine tumour
CC necrosis factor alpha (TNFalpha) convertase of the invention.
CC TNFalpha-convertase can proteolytically convert TNFalpha precursor to
CC mature TNFalpha. The convertase can be used to isolate novel compounds
CC capable of binding to it, which preferably inhibit its activity.
CC Inhibitors of the convertase are useful to treat a disease or condition
CC characterised by an elevated level of TNFalpha in the serum or tissues of
CC a mammal, e.g. systemic inflammatory response syndrome, reperfusion
CC injury, cardiovascular disease, infectious disease, obstetrical
CC disorders, gynaecological disorders, inflammatory disease, autoimmunity,
CC allergic disease, atopic disease, malignancy, transplant complication,
CC septic shock, cachexia, AIDS, graft osteoporosis, restenosis, psoriasis,
CC infarction (preferably due to an ischaemic event), rheumatoid arthritis,
CC macular degeneration, osteoarthritis or multiple sclerosis. The
CC TNFalpha-convertase inhibitors can be modified for use as ligands to
CC purify TNFalpha-convertase.
XX
SQ Sequence 2464 BP; 605 A; 630 C; 659 G; 570 T; 0 other:

Query Match 35.6%; Score 22.8; DB 18; Length 2464;
Best Local Similarity 79.4%; Pred. No. 20;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 27 cactagagctctgctgctgctgctcttcgcg 60
   | | | | | | | | | | | | | | | | | |
Db 2302 cccctagacctgctgctgctgctgcgcgcg 2335

```

Search completed: February 12, 2001, 21:27:43
 Job time: 3539 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2001, 20:11:54 ; Search time 1275.75 Seconds

(without alignments)
256,740 Million cell updates/sec

Title: US-09-600-602-1

Perfect score: 64
Sequence: 1 taagcccaagtgtcttaagc.....gcttgctctctcgcgtca 64

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_om: *
4: gb_ov: *
5: gb_ph: *
6: gb_pi1: *
7: gb_pi2: *
8: gb_pi3: *
9: gb_pi4: *
10: gb_pi5: *
11: gb_pi6: *
12: gb_pi7: *
13: gb_pi8: *
14: gb_pi9: *
15: gb_pi10: *
16: gb_pi11: *
17: gb_pi12: *
18: gb_pi13: *
19: gb_pi14: *
20: gb_pi15: *
21: gb_pi16: *
22: gb_pi17: *
23: gb_pi18: *
24: gb_pi19: *
25: gb_pi20: *
26: gb_pi21: *
27: gb_pi22: *
28: gb_pi23: *
29: gb_pi24: *
30: gb_pi25: *
31: gb_pi26: *
32: gb_pi27: *
33: gb_pi28: *
34: gb_pi29: *
35: gb_pi30: *
36: gb_pi31: *
37: gb_pi32: *
38: gb_pi33: *
39: gb_pi34: *
40: gb_pi35: *
41: gb_pi36: *
42: gb_pi37: *
43: gb_pi38: *

44: em_htg2: *
45: em_htg3: *
46: em_htg4: *
47: em_htg5: *
48: em_htg6: *
49: em_htg7: *
50: em_htg8: *
51: em_htg9: *
52: em_htg10: *
53: em_htg11: *
54: em_htg12: *
55: em_htg13: *
56: em_htg14: *
57: em_htg15: *
58: em_htg16: *
59: em_htg17: *
60: em_htg18: *
61: em_htg19: *
62: em_htg20: *
63: em_htg21: *
64: em_htg22: *
65: em_htg23: *
66: em_htg24: *
67: em_htg25: *
68: em_htg26: *
69: em_htg27: *
70: em_htg28: *
71: em_htg29: *
72: em_htg30: *
73: em_htg31: *
74: em_htg32: *
75: em_htg33: *
76: em_htg34: *
77: em_htg35: *
78: em_htg36: *
79: em_htg37: *
80: em_htg38: *
81: em_htg39: *
82: em_htg40: *
83: em_htg41: *
84: em_htg42: *
85: em_htg43: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	100.0	173	81	AR037064
2	64	100.0	183	81	AR037061
3	64	100.0	2799	81	AR005013
4	64	100.0	2799	81	AR037063
5	64	100.0	5871	6	AB001920
6	28	43.8	161234	48	CNS01DUE
7	26.4	41.2	87974	49	AC021283
8	26.4	41.2	162866	50	AC023277
9	26.4	41.2	173556	50	AC074388
10	26.4	41.2	207550	40	AC015977
11	26.4	41.2	215708	57	AC068757
12	26.2	40.9	6036	11	RATPDC
13	26	40.6	110000	84	HSX1_1
14	26	40.6	110000	84	HSX1_2
15	26	40.6	300050	78	HSX1A
16	26	40.6	340000	77	HS21C084
17	25.8	40.3	958	79	CNS01M4
18	25.6	40.0	76141	51	AC024393
19	25.6	40.0	106784	66	AL139250
20	25.6	40.0	168982	69	AL390994
21	25.4	39.7	21721	32	CEP16H9

```
22 25.4 39.7 172551 9 AC008518
C 23 25.4 39.7 176616 51 AC026427
C 24 25.4 39.7 186340 41 AC016684
25 25.2 39.4 3893 6 AF036763
26 25.2 39.4 5626 6 AF043332
C 27 25.2 39.4 111810 50 AC022139
C 28 25.2 39.4 147880 49 AC021945
29 25.2 39.4 152290 40 AC016107
C 30 25.2 39.4 169069 30 AC008871
C 31 25.2 39.4 179686 51 AC023987
C 32 25.2 39.4 186603 54 AC036155
C 33 25.2 39.4 192638 41 AC017063
34 25.2 39.4 206074 10 AC009484
35 25.2 39.1 41193 31 AC013257
C 36 25 39.1 116215 8 AC004130
C 37 25 39.1 172083 10 AC025614
C 38 25 39.1 219194 49 AC022061
C 39 24.8 38.8 370 11 MUSCMOSR2
C 40 24.8 38.8 539 11 MUSRCMOS
41 24.8 38.8 1331 81 E01225
42 24.8 38.8 2110 11 MMEFV1A
43 24.8 38.8 2110 81 A67510
44 24.8 38.8 88028 8 AC004938
C 45 24.8 38.8 97832 78 HSDJ46801
```

ALIGNMENTS

```
RESULT 1
LOCUS AR037064 173 bp DNA
DEFINITION Sequence 7 from patent US 5801016.
ACCESSION AR037064
VERSION AR037064.1 GI:5954920
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 173)
AUTHORS Morloka,S. and Ueki,J.
TITLE DNA fragment, recombinant vector containing the same and method for
JOURNAL expressing foreign genes using the same
FEATURES
source location/Qualifiers
1..173
BASE COUNT 21 a 48 c 47 g 57 t
ORIGIN
```

```
Query Match 100.0%; Score 64; DB 81; Length 173;
Best Local Similarity 100.0%; Pred. No. 8.7e-13;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 taagccagtgctgcttaagcgaactagagctcttcgcttcgtcttcctcgcg 60
|||||
Db 2 TAAACCCAGTGTCTTAGGCTAAGCGCAGTACGACTTCTGCTGCTTCTTCGCG 61
```

```
OY 61 ctca 64
||||
Db 62 CTCA 65
```

```
RESULT 2
LOCUS AR037061 183 bp DNA
DEFINITION Sequence 1 from patent US 5801016.
ACCESSION AR037061
VERSION AR037061.1 GI:5954917
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
```

```
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 183)
Morloka,S. and Ueki,J.
TITLE DNA fragment, recombinant vector containing the same and method for
JOURNAL expressing foreign genes using the same
FEATURES
source location/Qualifiers
1..183
BASE COUNT 23 a 51 c 51 g 58 t
ORIGIN
```

```
Query Match 100.0%; Score 64; DB 81; Length 183;
Best Local Similarity 100.0%; Pred. No. 8.7e-13;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 taagccagtgctgcttaagcgaactagagctcttcgcttcgtcttcctcgcg 60
|||||
Db 7 TAAACCCAGTGTCTTAGGCTAAGCGCAGTACGACTTCTGCTGCTTCTTCGCG 66
```

```
OY 61 ctca 64
||||
Db 67 CTCA 70
```

```
RESULT 3
LOCUS AR005013 2799 bp DNA
DEFINITION Sequence 5 from patent US 5747327.
ACCESSION AR005013
VERSION AR005013.1 GI:3965892
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2799)
AUTHORS Ueki,J. and Morloka,S.
TITLE Phospholipase D gene originated from plant
JOURNAL Patent: US 5747327-A 5 05-MAY-1998;
FEATURES
source location/Qualifiers
1..2799
BASE COUNT 692 a 709 c 609 g 789 t
ORIGIN
```

```
Query Match 100.0%; Score 64; DB 81; Length 2799;
Best Local Similarity 100.0%; Pred. No. 9.6e-13;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 taagccagtgctgcttaagcgaactagagctcttcgcttcgtcttcctcgcg 60
|||||
Db 1667 TAAACCCAGTGTCTTAGGCTAAGCGCAGTACGACTTCTGCTGCTTCTTCGCG 1726
```

```
OY 61 ctca 64
||||
Db 1727 CTCA 1730
```

```
RESULT 4
LOCUS AR037063 2799 bp DNA
DEFINITION Sequence 4 from patent US 5801016.
ACCESSION AR037063
VERSION AR037063.1 GI:5954919
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2799)
AUTHORS Morloka,S. and Ueki,J.
TITLE DNA fragment, recombinant vector containing the same and method for
```


expressing foreign genes using the same
Patent: US 5801016-A 4 01-Sep-1998;
Location/Qualifiers
1. .2799
/organism="unknown"
BASE COUNT 692 a 709 c 609 g 789 t
ORIGIN

Query Match 100.0%; Score 64; DB 81; Length 2799;
Best Local Similarity 100.0%; Pred. No. 9.6e-13;
Matches .64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taagccagtgctgaagcgaagcagcttcgtcgtctctccg 60
|||||
Db 1667 TAAGCCAGTGTGCTTAGCCTAAGCGCAGCTTCTGCTGCTTCTCTCG 1726

QY 61 ctca 64
|||||
Db 1727 CTCA 1730

RESULT 5
AB001920 5871 bp DNA PLN 14-APR-2000
LOCUS
DEFINITION Oryza sativa DNA for phospholipase D, complete cds.
ACCESSION AB001920
VERSION AB001920.1 GI:1902902
KEYWORDS phospholipase D.
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE
AUTHORS Ueki,J., Morioka,S., Komari,T. and Kumashiro,T.
TITLE 1 (bases 1522 to 1665; 1839 to 1983; 2524 to 4420)
Purification and characterization of phospholipase D (PLD) from rice (Oryza sativa L.) and cloning of cDNA for PLD from rice and maize (Zea mays L.)
JOURNAL Plant Cell Physiol. 36 (5), 903-914 (1995)
MEDLINE 96012933
REFERENCE
AUTHORS Morioka,S., Ueki,J. and Komari,T.
TITLE 2 (bases 1 to 5871)
Characterization of two distinctive genomic clones (Accession Nos. AB001919 and AB001920) for phospholipase D from rice (PER97-076)
JOURNAL Plant Physiol. 114, 396 (1997)
REFERENCE
AUTHORS Ueki,J.
TITLE 3 (bases 1 to 5871)
Direct Submission
Submitted (11-OCT-1995) to the DDBJ/EMBL/Genbank databases. Jun Ueki, Japan Tobacco Inc., Plant Breeding and Genetics Research Lab; 700 Higashihara, Iwata, Shizuoka 438-0802, Japan (E-mail:jun.ueki@pbgrl.jti.co.jp, Tel:81-538-32-7111, Fax:81-538-32-8700)
FEATURES
source Location/Qualifiers
1. .5871
/organism="Oryza sativa"
/cultivar="Koshihikari"
/db_xref="taxon:4530"
/tissue_type="leaf"
join(1522. .1665,1839. .1983,2524. .4420,4899. .5702)
/citation=[1]
1522. .1665
/number=1
1666. .1838
/number=1
1839. .1983
/number=2
join(1876. .1983,2524. .4420,4899. .5332)
/EC_number="3.1.4.4"
/citation=[1]
/codon_start=1
/product="phospholipase D"
/protein_id="BAA19467.1"

/db_xref="GI:1902903"
/translation="MAQMLLHGLTHATIFEAASLSNPHRASGAPKIRKFEVGEIDET
VGWKGATIKVYSTIDLEKARVGRTRMTNPIINRWESPHYICAHASNVIFVKID
NIGATNIGRAYLIPVQELLNGEELIDCDNNRESVGSKIHKVKQYDVSQDM
ARGVRSKTYGVPVPTFFSQGGCAVTILKQAHAPDNIPKIPLADKNYEPHRCEPI
FDATSNQAHLLVITGMSVYETILVROSNRPKPGDVTLLELKKASEVRLVLMV
DIDRTSGLKLRDGLMAHDETEHYFPGSDVNCVLCRNDDSGSIVQDLSISTMFPH
HOKIVVVDHELPLNGSGOORIVSFVGGDLDCDXYDYQYHSLFRTLDSTHDDPHON
FATASIKKGGPREPMHDIHSLRLEGIAMVLYNFEQRRKGGKDLLQLRSDTI
PPSPVMPEDPREMTNVLFRSIDGGAFFGEPDPEAKAGIVSGKQIIDRSIOAY
IHARRAKNTYITENOYFLUSSYAKREPIKPEDTGLHLIPKELAKVSKIEAGER
FTYIVVVPMPPEGVESGVQALLDWQRRTMEMYTDITLALQKKEIAPKQYLTFF
CLGNREVKQAGEYQPEQPRADNDYSPAQEARREMYIVHTKMLVDEYIITGSANIN
QSRMDGARDESIAMGYOPHYLATROPARCOIHGRMALWYELHGMIDVDFORPESLE
CVOKVNRIAEKYMDMYSDDLQDLPGLHLSYPIGVASDGVTELPQMEYFPDTRAV
LGAKSDYMPILITS"
1984. .2523
/number=2
2524. .4420
/number=3
4421. .4898
/number=3
4899. .5702
/number=4
BASE COUNT 1488 a 1363 c 1355 g 1665 t
ORIGIN

Query Match 100.0%; Score 64; DB 6; Length 5871;
Best Local Similarity 100.0%; Pred. No. 9.9e-13;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taagccagtgctgaagcgaagcagcttcgtcgtctctccg 60
|||||
Db 1667 TAAGCCAGTGTGCTTAGCCTAAGCGCAGCTTCTGCTGCTTCTCTCG 1726

QY 61 ctca 64
|||||
Db 1727 CTCA 1730

RESULT 6
CNS01DUE/c DNA PRI 25-MAY-2000
LOCUS
DEFINITION Human chromosome 14 DNA sequence *** IN PROGRESS *** BAC C-2560C21
of library Caltech-D from chromosome 14 of Homo sapiens (human),
complete sequence.
ACCESSION AL133241
VERSION AL133241.3 GI:7630038
KEYWORDS HTG; HTGS_DRAFT.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 161234)
Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2000) to the EMBL/Genbank/DDBJ databases
COMMENT On Apr 20, 2000 this sequence version replaced gi:6624578.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc. . . even if efforts are made to eliminate these
contaminating sequences. The following BAC sequence is oriented
from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-61A17
Downstream BAC (overlapping the SP6 end) : C-2379B19

Overall quality chart :
Range : bases
0


```

* 25957 26824: contig of 868 bp in length
* 26825 26924: gap of 100 bp
* 26925 27814: contig of 890 bp in length
* 27815 27914: gap of 100 bp
* 27915 28785: contig of 871 bp in length
* 28786 28885: gap of 100 bp
* 28886 29772: contig of 887 bp in length
* 29773 29872: gap of 100 bp
* 29873 30730: contig of 858 bp in length
* 30731 30830: gap of 100 bp
* 30831 31680: contig of 850 bp in length
* 31681 31780: gap of 100 bp
* 31781 32654: contig of 874 bp in length
* 32655 32754: gap of 100 bp
* 32755 33621: contig of 867 bp in length
* 33622 33721: gap of 100 bp
* 33722 34595: contig of 874 bp in length
* 34596 34695: gap of 100 bp
* 34696 35573: contig of 878 bp in length
* 35574 35673: gap of 100 bp
* 35674 36571: contig of 898 bp in length
* 36572 36671: gap of 100 bp
* 36672 37544: contig of 873 bp in length
* 37545 37644: gap of 100 bp
* 37645 38522: contig of 878 bp in length
* 38523 38622: gap of 100 bp
* 38623 39484: contig of 862 bp in length
* 39485 39584: gap of 100 bp
* 39585 40468: contig of 884 bp in length
* 40469 40568: gap of 100 bp
* 40569 41434: contig of 866 bp in length
* 41435 41534: gap of 100 bp
* 41535 42393: contig of 859 bp in length
* 42394 42493: gap of 100 bp
* 42494 43380: contig of 887 bp in length
* 43381 43480: gap of 100 bp
* 43481 44375: contig of 895 bp in length
* 44376 44475: gap of 100 bp
* 44476 45341: contig of 866 bp in length
* 45342 45441: gap of 100 bp
* 45442 46302: contig of 861 bp in length
* 46303 46402: gap of 100 bp
* 46403 47300: contig of 898 bp in length
* 47301 47400: gap of 100 bp
* 47401 48288: contig of 888 bp in length
* 48289 48388: gap of 100 bp
* 48389 49261: contig of 873 bp in length
* 49262 49361: gap of 100 bp
* 49362 50236: contig of 875 bp in length
* 50237 50336: gap of 100 bp
* 50337 51222: contig of 886 bp in length
* 51223 51322: gap of 100 bp
* 51323 52185: contig of 863 bp in length
* 52186 52285: gap of 100 bp
* 52286 53149: contig of 864 bp in length
* 53150 53249: gap of 100 bp
* 53250 54086: contig of 837 bp in length
* 54087 54186: gap of 100 bp
* 54187 55046: contig of 860 bp in length
* 55047 55146: gap of 100 bp
* 55147 55995: contig of 849 bp in length
* 55996 56095: gap of 100 bp
* 56096 56975: contig of 880 bp in length
* 56976 57075: gap of 100 bp
* 57076 57960: contig of 885 bp in length
* 57961 58060: gap of 100 bp
* 58061 58938: contig of 878 bp in length
* 58939 59038: gap of 100 bp
* 59039 59917: contig of 879 bp in length
* 59918 60017: gap of 100 bp
* 60018 60891: contig of 874 bp in length
* 60892 60991: gap of 100 bp
* 60992 61867: contig of 876 bp in length

```

```

Query Match 41.2%; Score 26.4; DB 49; Length 87974;
Best Local Similarity 69.2%; Pred. No. 39;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 3 agccagatgtgttagtgtaagcagcactagagctctctgtctgtcttct 54
DB 66387 AGTGCAGTGGCCCAATCTTGCTCATCTGAGACCTGTGCTCCGCGCTTCT 66438

```

```

RESULT 8
AC023277/c AC023277 162866 bp DNA HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 4 clone RP11-438E5, WORKING DRAFT SEQUENCE,
DEFINITION 27 unordered clones.
ACCESSION AC023277
VERSION AC023277.3 GI:7712273
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eularchia; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 162866)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 162866)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On May 6, 2000 this sequence version replaced gi:7024025.
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0438E05
----- Summary Statistics -----
Sequencing vector: M13; %
Chemistry: Dye-Primer ET; % of reads
Chemistry: Dye-terminator Big Dye; % of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 146574 bases at least Q40
Consensus quality: 151916 bases at least Q30
Consensus quality: 154698 bases at least Q20
Insert size: 160266; sum-of-ctrls
Quality coverage: 3.67 in Q20 bases; agarose-fp
Quality coverage: 3.76 in Q20 bases; sum-of-ctrls

```

* NOTE: This is a 'working draft' sequence. It currently consists of 27 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 173556)
Waterston,R.H.
Direct Submission
Submitted (30-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 4, 2000 this sequence version replaced g1:9587428.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Project name: H.NH0507P19
Center project name: Summary Statistics

Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-Primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 158826 bases at least Q40
Consensus quality: 163318 bases at least Q40
Consensus quality: 165604 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 171556; sum-of-contigs
Quality coverage: 4.60 in Q20 bases; agarose-fp
Quality coverage: 4.61 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2144: contig of 2144 bp in length
* 2145 2244: gap of unknown length
* 2245 4052: contig of 1808 bp in length
* 4053 4152: gap of unknown length
* 4153 6306: contig of 2154 bp in length
* 6307 6406: gap of unknown length
* 6407 8444: contig of 2038 bp in length
* 8445 8544: gap of unknown length
* 8545 11626: contig of 3082 bp in length
* 11627 11726: gap of unknown length
* 11727 15823: contig of 4097 bp in length
* 15824 15923: gap of unknown length
* 15924 20350: contig of 4427 bp in length
* 20351 20450: gap of unknown length
* 20451 25976: contig of 5526 bp in length
* 25977 26076: gap of unknown length
* 26077 31313: contig of 5237 bp in length
* 31314 31413: gap of unknown length
* 31414 38673: contig of 7260 bp in length
* 38674 38773: gap of unknown length
* 38774 45342: contig of 6569 bp in length
* 45343 45442: gap of unknown length
* 45443 53428: contig of 7986 bp in length
* 53429 53529: gap of unknown length
* 53530 60068: contig of 6540 bp in length
* 60069 60169: gap of unknown length
* 60170 68196: contig of 8028 bp in length
* 68197 68296: gap of unknown length
* 68297 76091: contig of 7795 bp in length
* 76092 76191: gap of unknown length
* 76192 87826: contig of 11635 bp in length
* 87827 87926: gap of unknown length
* 87927 99071: contig of 11145 bp in length
* 99072 99171: gap of unknown length
* 99172 115184: contig of 16013 bp in length

* 115185 115284: gap of unknown length
* 115285 132703: contig of 17419 bp in length
* 132704 132803: gap of unknown length
* 132804 132804: contig of 21261 bp in length
* 154065 154164: gap of unknown length
* 154165 173556: contig of 19392 bp in length.
Location/Qualifiers
1. 173556

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-507P19"
1. 2144
misc_feature
/note="assembly_name:Contig11"
2245. 4052
/note="assembly_name:Contig12"
4153. 6306
misc_feature
/note="assembly_name:Contig13"
6407. 8444
/note="assembly_name:Contig14"
8545. 11626
misc_feature
/note="assembly_name:Contig15"
11727. 15823
/note="assembly_name:Contig16"
15924. 20350
misc_feature
/note="assembly_name:Contig17"
20451. 25976
/note="assembly_name:Contig18"
26077. 31313
misc_feature
/note="assembly_name:Contig19"
31414. 38673
/note="assembly_name:Contig20"
38774. 45342
misc_feature
/note="assembly_name:Contig21"
45443. 53428
/note="assembly_name:Contig22"
53529. 60068
misc_feature
/note="assembly_name:Contig23"
60169. 68196
/note="assembly_name:Contig24"
68297. 76091
/note="assembly_name:Contig25"
76192. 87826
misc_feature
/note="assembly_name:Contig26"
87927. 99071
misc_feature
/note="assembly_name:Contig27"
99172. 115184
/note="assembly_name:Contig28"
115285. 132703
misc_feature
/note="assembly_name:Contig29"
clone_end:sp6
vector_side:left"
132804. 154064
/note="assembly_name:Contig30"
154165. 173556
misc_feature
/note="assembly_name:Contig31"
BASE COUNT 42375 a 43249 c 43427 g 42464 t 2041 others
ORIGIN

Query Match 41.2% Score 26.4; DB 59; Length 173556;
Best Local Similarity 69.2%; Pred. No. 40; Mismatches 16; Indels 0; Gaps 0;

Oy 12 tgcctagcctaagcgcactagagctctgctgcctctctccgcgc 63

||||| ||||| | | | | | ||||| ||| ||||| ||||| | |
DB 140974 TGCTGTGGCTATGCGGCGCAAGTCTCTTGCTTACTTCTCTCTCCCTCC 140923

RESULT 10
AC015977/c

LOCUS AC015977 .207550 bp DNA HTG 17-AUG-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-527I18, WORKING DRAFT
AC015977
SEQUENCE: 23 unordered pieces.
AC015977
AC015977.7 GI:9838247
VERSION
KEYWORDS
SOURCE HTG: HTGS_PHASE1; HTGS_DRAFT.
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 207550)
Waterston, R.H.
The sequence of Homo sapiens clone
unpublished
2 (bases 1 to 207550)
Waterston, R.H.
Direct Submission
Submitted (17-NOV-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 17, 2000 this sequence version replaced g1:9211469.

51986 52085: gap of unknown length
* 52086 64048: contig of 11963 bp in length
* 64049 64148: gap of unknown length
* 64149 73711: contig of 9563 bp in length
* 73712 73811: gap of unknown length
* 73812 81475: contig of 7664 bp in length
* 81476 81575: gap of unknown length
* 81576 91789: contig of 10114 bp in length
* 91790 102726: contig of 10937 bp in length
* 102727 102826: gap of unknown length
* 102827 114269: gap of 11443 bp in length
* 114270 114369: gap of unknown length
* 114370 130207: contig of 15838 bp in length
* 130208 130307: gap of unknown length
* 130308 148001: contig of 17694 bp in length
* 148002 148101: gap of unknown length
* 148102 166521: contig of 18420 bp in length
* 166522 166621: gap of unknown length
* 166622 186379: contig of 19758 bp in length
* 186380 207550: gap of unknown length
* 186480 207550: contig of 21071 bp in length.

FEATURES

source

1. .207550
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-527I18"
1. .2257
/note="assembly_name:Contig14"
2358. .4468
/note="assembly_name:Contig15"
4569. .7877
/note="assembly_name:Contig16"
7978. .11086
/note="assembly_name:Contig17
clone_end:SP6
vector_side:left"
1187. .14728
/note="assembly_name:Contig18"
14829. .19922
/note="assembly_name:Contig19"
20023. .25092
/note="assembly_name:Contig20"
25193. .30468
/note="assembly_name:Contig21"
30569. .34509
/note="assembly_name:Contig22"
34610. .40469
/note="assembly_name:Contig23"
40570. .46884
/note="assembly_name:Contig24"
46985. .51985
/note="assembly_name:Contig25"
52086. .64048
/note="assembly_name:Contig26"
64149. .73711
/note="assembly_name:Contig27"
73812. .81475
/note="assembly_name:Contig28"
81576. .91689
/note="assembly_name:Contig29"
91790. .102726
/note="assembly_name:Contig30"
102827. .114269
/note="assembly_name:Contig31"
114370. .130207
/note="assembly_name:Contig32"
130308. .148001
/note="assembly_name:Contig33"
148102. .166521
/note="assembly_name:Contig34"
166622. .186379

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Project Information
Center project name: H_NH0527I18
----- Summary Statistics -----
Sequencing vector: MJ3. 658
Sequencing vector: plasmid; 358
Chemistry: Dye-primer ET; 65% of reads
Chemistry: Dye-terminator Big Dye; 35% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 193553 bases at least Q40
Consensus quality: 198962 bases at least Q30
Consensus quality: 201706 bases at least Q20
Insert size: 201000; agarose-fp
Insert size: 207759; sum-of-contigs
Quality coverage: 4.48 in Q20 bases; sum-of-contigs
Quality coverage: 4.40 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2257: contig of 2257 bp in length
2358 2357: gap of unknown length
2358 4468: contig of 2111 bp in length
4469 4568: gap of unknown length
4569 7877: contig of 3309 bp in length
7878 7977: gap of unknown length
7978 11086: contig of 3109 bp in length
11087 11186: gap of unknown length
11187 14728: contig of 3542 bp in length
14729 14828: gap of unknown length
14829 19922: contig of 5094 bp in length
19923 20022: gap of unknown length
20023 25092: contig of 5070 bp in length
25093 25192: gap of unknown length
25193 30468: contig of 5276 bp in length
30469 30568: gap of unknown length
30569 34509: contig of 3941 bp in length
34510 34609: gap of unknown length
34610 40469: contig of 5860 bp in length
40470 40569: gap of unknown length
40570 46884: contig of 6315 bp in length
46885 46984: gap of unknown length
46985 51985: contig of 5001 bp in length

THIS PAGE BLANK (USPTO)

OY 61 ctca 64
|||||
Db 62 CTCA 65

RESULT 2

US-08-750-007-1
; Sequence 1, Application US/08750007
; Patent No. 5801016
; GENERAL INFORMATION:
; APPLICANT: MORIOKA, SHINJI
; APPLICANT: UEKI, JUN
; TITLE OF INVENTION: DNA FRAGMENT, RECOMBINANT VECTOR
; TITLE OF INVENTION: CONTAINING THE SAME AND METHOD FOR EXPRESSING FOREIGN
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750.007
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0760-221P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 183 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-750-007-1

Query Match 100.0%; Score 64; DB 1; Length 183;
Best Local Similarity 100.0%; Pred. No. 1.4e-15;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 taagccagtgcttgctgaagcactagagcttctgctgcttcttcgcg 60
|||||
Db 7 TAACCCAGTGCTGTAGGCTAAGCGCAGCTTCTGCTTCTTCGCG 66

OY 61 ctca 64
|||||
Db 67 CTCA 70

RESULT 3

US-08-446-794A-5
; Sequence 5, Application US/08446794A
; Patent No. 5747327
; GENERAL INFORMATION:
; APPLICANT: UEKI, JUN
; APPLICANT: MORIOKA, SHINJI
; TITLE OF INVENTION: PHOSPHOLIPASE D GENE ORIGINATED FROM
; NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: P.O. BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446.794A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0760-0203P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2799 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1876..1983
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2524..2799
; US-08-446-794A-5

Query Match 100.0%; Score 64; DB 1; Length 2799;
Best Local Similarity 100.0%; Pred. No. 3e-15;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 taagccagtgcttgctgaagcactagagcttctgctgcttcttcgcg 60
|||||
Db 1667 TAACCCAGTGCTGTAGGCTAAGCGCAGCTTCTGCTTCTTCGCG 1726

OY 61 ctca 64
|||||
Db 1727 CTCA 1730

RESULT 4

US-08-750-007-4
; Sequence 4, Application US/08750007
; Patent No. 5801016
; GENERAL INFORMATION:
; APPLICANT: MORIOKA, SHINJI
; APPLICANT: UEKI, JUN
; TITLE OF INVENTION: DNA FRAGMENT, RECOMBINANT VECTOR
; TITLE OF INVENTION: CONTAINING THE SAME AND METHOD FOR EXPRESSING FOREIGN
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

Gencore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2001, 19:56:04 ; Search time 1923.53 Seconds
(without alignments)
233.154 Million cell updates/sec

Title: US-09-600-602-1

Perfect score: 64
Sequence: 1 taagccagctgtccttaagc.....gcttgcctctctccgcctca 64

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: em_estda.*
42: em_estfun.*
43: em_esthum1.*

44: em_esthum2.*
45: em_esthum3.*
46: em_esthum4.*
47: em_esthum5.*
48: em_esthum6.*
49: em_esthum7.*
50: em_esthum8.*
51: em_esthum9.*
52: em_esthum10.*
53: em_esthum11.*
54: em_esthum12.*
55: em_esthum13.*
56: em_esthum14.*
57: em_esthum15.*
58: em_esthum16.*
59: em_esthum17.*
60: em_esthum18.*
61: em_esthum19.*
62: em_esthum20.*
63: em_esthum21.*
64: em_esthum22.*
65: em_esthum23.*
66: em_esthum24.*
67: em_esthum25.*
68: em_esthum26.*
69: em_esthum27.*
70: em_esthum28.*
71: em_esthum29.*
72: em_esthum30.*
73: em_esthum31.*
74: em_esthum32.*
75: em_esthum33.*
76: em_esthum34.*
77: em_esthum35.*
78: em_esthum36.*
79: em_esthum37.*
80: em_esthum38.*
81: em_esthum39.*
82: em_esthum40.*
83: em_esthum41.*
84: em_esthum42.*
85: em_esthum43.*
86: em_esthum44.*
87: em_esthum45.*
88: em_esthum46.*
89: em_esthum47.*
90: em_esthum48.*
91: em_esthum49.*
92: em_esthum50.*
93: em_esthum51.*
94: em_esthum52.*
95: em_esthum53.*
96: em_esthum54.*
97: em_esthum55.*
98: em_esthum56.*
99: em_esthum57.*
100: em_esthum58.*
101: em_esthum59.*
102: em_esthum60.*
103: em_esthum61.*
104: em_esthum62.*
105: em_esthum63.*
106: em_esthum64.*
107: em_esthum65.*
108: em_esthum66.*
109: em_esthum67.*
110: em_esthum68.*
111: em_esthum69.*
112: em_esthum70.*
113: em_esthum71.*
114: em_esthum72.*
115: em_esthum73.*
116: em_esthum74.*

117: em_estp16:*
118: em_estp17:*
119: em_estp18:*
120: em_estp14:*
121: em_estp15:*
122: em_estp16:*
123: em_estp17:*
124: em_estp18:*
125: em_estp19:*
126: gb_est58:*
127: gb_est59:*
128: gb_est60:*
129: gb_est61:*
130: gb_est62:*
131: gb_est63:*
132: gb_est64:*
133: gb_est65:*
134: gb_est66:*
135: gb_est67:*
136: gb_est68:*
137: gb_est69:*
138: gb_est70:*
139: gb_est71:*
140: gb_est72:*
141: gb_est73:*
142: gb_est74:*
143: gb_est75:*
144: gb_est76:*
145: gb_est77:*
146: gb_est78:*
147: gb_est79:*
148: gb_est80:*
149: gb_est81:*
150: gb_est82:*
151: gb_est83:*
152: gb_est84:*
153: gb_est85:*
154: gb_est86:*
155: gb_est87:*
156: gb_est88:*
157: gb_est89:*
158: gb_est90:*
159: gb_est91:*
160: gb_est92:*
161: gb_est93:*
162: gb_est94:*
163: gb_est95:*
164: gb_est96:*
165: gb_est97:*
166: gb_est98:*
167: gb_est99:*
168: gb_est100:*
169: gb_est101:*
170: gb_est102:*
171: gb_est103:*
172: gb_est104:*
173: gb_est105:*
174: gb_est106:*
175: gb_est107:*
176: gb_est108:*
177: gb_est109:*
178: gb_est110:*
179: gb_est111:*
180: gb_est112:*
181: gb_est113:*
182: gb_est114:*
183: gb_est115:*
184: gb_est116:*
185: gb_est117:*
186: gb_est118:*
187: gb_est119:*
188: gb_est120:*
189: gb_est121:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	31	48.4	356	95	AW763066	AW763066 ur60d05.y
2	31	48.4	474	88	AW259340	AW259340 um92c02.y
3	31	48.4	485	147	W79969	W79969 me90d09.r1
4	31	48.4	533	20	A1425798	A1425798 me90d09.y
5	31	48.4	540	14	AA339633	AA339633 vz93h10.r
6	31	48.4	829	106	BE285743	BE285743 601096312
7	29.4	45.9	312	129	BB381871	BB381871 BB381871
8	27.4	42.8	746	107	BE415895	BE415895 MUC002.B0
9	26.2	40.9	221	172	AQ933849	AQ933849 RPT-23-2
10	26.2	40.9	281	125	BB410001	BB410001 BB410001
11	26.2	40.9	296	130	BB410925	BB410925 BB410925
12	26.2	40.9	302	103	BB226614	BB226614 BB226614
13	25.8	40.3	465	182	AZ496732	AZ496732 1M033317
14	25.8	40.3	664	94	AW574773	AW574773 874001D08
15	25.6	40.0	764	175	AZ118240	AZ118240 RPT-23-4
16	25.6	40.0	1052	29	AU050386	AU050386 AU050386
17	25.2	39.4	300	126	BB263403	BB263403 BB263403
18	25.2	39.4	328	132	BB497860	BB497860 BB497860
19	25	39.1	281	104	BE128709	BE128709 DEPA2459
20	25	39.1	283	139	C23376	C23376 C23376 Japa
21	25	39.1	378	40	AM126420	AM126420 614071C05
22	25	39.1	494	29	AU084524	AU084524 AU084524
23	25	39.1	515	21	A1503026	A1503026 VM75d11.x
24	25	39.1	526	157	AQ459497	AQ459497 HS-5085_A
25	25	39.1	540	108	BE469225	BE469225 IPHdK0119
26	25	39.1	543	108	BE469051	BE469051 IPHdK0316
27	25	39.1	545	175	AZ102103	AZ102103 RPT-23-1
28	25	39.1	548	181	AZ401292	AZ401292 IM0167L20
29	25	39.1	559	39	AW052990	AW052990 614077E04
30	25	39.1	580	88	AW259420	AW259420 up29d02.x
31	25	39.1	602	38	AW013436	AW013436 PC031KS M
32	25	39.1	626	39	AW053148	AW053148 614032H06
33	25	39.1	634	24	A1712055	A1712055 614004G06
34	25	39.1	860	97	AW983363	AW983363 HVSME9001
35	24.8	38.8	179	93	AW607959	AW607959 CM3-HT051
36	24.8	38.8	182	10	AA653836	AA653836 ns94h12.s
37	24.8	38.8	328	13	AA920859	AA920859 vY83f06.r
38	24.8	38.8	331	175	AZ111044	AZ111044 RPT-23-4
39	24.8	38.8	458	18	A1727262	A1727262 UK06C02.y
40	24.8	38.8	524	174	AZ079555	AZ079555 RPT-23-3
41	24.8	38.8	575	173	AZ086580	AZ086580 RPT-23-2
42	24.8	38.8	685	173	AQ098548	AQ098548 RPT-23-3
43	24.8	38.8	852	168	AQ745431	AQ745431 HS-2278_A
44	24.6	38.4	295	156	AQ390792	AQ390792 CITR1-EI-
45	24.6	38.4	325	184	B79194	B79194 CTR78SK-30

ALIGNMENTS

RESULT 1
LOCUS AW763066 356 bp mRNA
DEFINITION ur60d05.y1 NCICGAP_Mam3 Mus musculus cDNA clone IMAGE:3154665 5',
ACCESION AW763066
VERSION AW763066.1 GI:7695001
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 356)
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/tresources.shtml
MGI:1057421
Seq primer: -40RP from Gibco
High quality sequence stop: 332.
Location/Qualifiers
1. 356
/organism="Mus musculus"
/strain="129 - C57/B6 - FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3154665"
/clone_lib="NCI-CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site.1: SalI;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."
BASE COUNT 93 a 57 c 89 g 117 t
ORIGIN
Query Match 48.4%; Score 31; DB 95; Length 356;
Best Local Similarity 72.7%; Pred. No. 1;
Matches 40; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
OY 5 cccagtgtgcttaagcactagaagcttctgtcgtcttcttc 59
|||||
Db 269 CCCAGTGTGCTGTACTACTACTACAGCTCTCAGCTGCAGGCTTCTCTCC 215
|||||
RESULT 2
LOCUS AM259340 474 bp mRNA EST 23-DEC-1999
DEFINITION um92c02.y1 Sugano mouse kidney mk1a Mus musculus cDNA clone
IMAGE:2333226 5', mRNA sequence.
ACCESSION AM259340
VERSION AM259340.1 GI:6632321
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 474)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Sallier,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1011430
Seq primer: custom primer used
High quality sequence stop: 466.
Location/Qualifiers
1. 474
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2332226"
/clone_lib="Sugano mouse kidney mk1a"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pME18S-FL3; Site.1: DraIII
(CACGCTG); Site.2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCTTTTATTTTATTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCTGCTCTAAACCTCGG and 3' end
primer CGACCTCAGCTCGACACA."
BASE COUNT 133 a 70 c 112 g 156 t 3 others
ORIGIN
Query Match 48.4%; Score 31; DB 88; Length 474;
Best Local Similarity 72.7%; Pred. No. 1.1;
Matches 40; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
OY 5 cccagtgtgcttaagcactagaagcttctgtcgtcttcttc 59
|||||
Db 321 CCCAGTGTGCTGTACTACTACTACAGCTCTCAGCTGCAGGCTTCTCTCC 267
|||||
RESULT 3
LOCUS W79969 485 bp mRNA EST 25-JUN-1996
DEFINITION me90d09.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone IMAGE:402833 5', mRNA sequence.
ACCESSION W79969
VERSION W79969.1 GI:1391071
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 485)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:246601
Seq primer: EPrPrimer
High quality sequence stop: 346.
Location/Qualifiers
1. 485
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="402833"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="vector: pT73d-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACCATCTGACAGTGGAGCGCGGAGAAATTTTCTTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT 139 a 94 c 109 g 143 t
ORIGIN

Query Match 48.4%; Score 31; DB 147; Length 485;
Best Local Similarity 72.7%; Pred. No. 1.1;
Matches 40; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 5 cccagtgcttagcgaagcagcctgcttctgctgcttctctcc 59
|||||
Db 98 CCCAGTGTCTGCTACCTTACTTACTACGCTTCACGCTTCCTTC 44

RESULT 4
A1425798 533 bp mRNA EST 15-MAR-2000
LOCUS me90609.y1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
DEFINITION clone IMAGE:402833 5', mRNA sequence.
ACCESSION A1425798
VERSION A1425798.1 GI:4271729
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 533)
AUTHORS Marr,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Pearson
B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Rittler
E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
CONTACT: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

TITLE
JOURNAL
COMMENT
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This read is a RESSEQUENCE of a previously sequenced mouse clone
MGI:246601
Seq primer: -40RP from Gibco

High quality sequence stop: 478
POLYA-No.
Location/Qualifiers
1. 533
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="402833"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="vector: pT73d-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACCATCTGACAGTGGAGCGCGGAGAAATTTTCTTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT 152 a 106 c 121 g 153 t 1 others
ORIGIN

Query Match 48.4%; Score 31; DB 20; Length 533;
Best Local Similarity 72.7%; Pred. No. 1.1;
Matches 40; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 5 cccagtgcttagcgaagcagcctgcttctgctgcttctctcc 59
|||||
Db 112 CCCAGTGTCTGCTACCTTACTTACTACGCTTCACGCTTCCTTC 58

RESULT 5
AA939633 540 bp mRNA EST 01-MAY-1998
LOCUS v293h10.r1 Soares_thymus_2NDMT Mus musculus cDNA clone
DEFINITION IMAGE:1344835 5', mRNA sequence.
ACCESSION AA939633
VERSION AA939633.1 GI:3100410
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 540)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
CONTACT: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

TITLE
JOURNAL
COMMENT
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 496.
Location/Qualifiers
1. 540
/organism="Mus musculus"

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1344835"
/clone_id="Soares_thymus_2NBWT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dn) primer [5',
TGTTACCAATCTGAAGTGGAGGCGCGGTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fátima Bonaldo."

```

Query Match	48.4%	Score	31	DB	14	length	540
Best Local Similarity	72.7%	Pred. No.	1.1				
Matches	40	Conservative	0	Mismatches	15	Indels	0
						Gaps	0

QY 5 cccagctgtgcttagcgacgacactagagctctctgtcgcctgtctctctccc 59
|||||
179 cccagctgtgcttagccttaaccttctacagctcttcacgtggcagggcttcttc 125
|||||

RESULT	6
BE285743/c	
LOCUS	BE285743
DEFINITION	829 bp mRNA
	601096312p1 NC1_CGAP_Mam5 Mus musculus CDNA clone IMAGE:3490970 5' , mRNA sequence.
	13-JUL-2000

VERSION	BE285743.1	GI:9163574
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE	1 (bases 1 to 829)
AUTHORS	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

Email: Robert.L.Strausberg@nlh.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA library Preparation: Life Technologies, Inc.
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.jnl.gov>
Plate: LLAH8534 row: g column: 03
High quality sequence stop: 646.

FEATURES	Location/Qualifiers
source	1. .829

```

/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/collection="IMAGE:3490970"
/clone_id="NCI_CGAP_Mem5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators: Providing samples: Lothar Hennighausen/Robin Humphreys,

```

	NIH ^a			
BASE COUNT	207 a	164 c	191 g	267 t
ORIGIN				

Query Match	48.4%	Score 31	DB 106	Length 829
Best Local Similarity	72.7%	Pred. No.	1.1	
Matches 40	Conservative 0	Mismatches 15	Indels 0	Gaps 0

Oy 5 ccacagtgcctcagcgcaagcactgagctcttcgtcttccccc 59
| | | | |
Db 261 cccagtgctgctgacctaacctttactacagctctcagctggcaggccttcttcttc 207
| | | | |

RESULT	7
BB381871/c	
LOCUS	
DEFINITION	
BB381871	312 bp mRNA
BB381871	Riken full-length enriched, 0 day neonate cerebellum Mus
musculus cdna C230016E14.3,	mRNA sequence.

ACCESSION	BB381871
VERSION	BB381871.1
	GI:9101601

SOURCE house mouse.

ORGANISM

REFERENCE 1 (bases 1 to 312)

AUTHORS

1 (bases 1 to 312)
 Komono H, Aizawa K, Akahira S, Akiyama J, Arahawa T, Carninci P, Endo T, Fukuda S, Fukunishi Y, Hata A, Hayatsu N, Hirozane T, Horii F, Ishii Y, Ishikawa J, Ishikawa T, Itoh M, Izawa M, Kadota K, Kagawa I, Kai C, Kawaj J, Kikuchi N, Kiyosawa H, Kojima Y, Kondo S, Koya S, Kurihara C, Kusabe M, Matsuyama T, Miki R, Mizuno Y, Nakamura M, Oda H, Okazaki Y, Ono T, Owa C, Salto H, Sakai C, Sato K, Shibata K, Shibata Y, Shigemoto Y, Shinagawa A, Shiraki T, Sogabe Y, Sugahara Y, Suzuki H, Suzuki H, Tagawa A, Takahashi F, Tomonaga N, Toy T, Tsunoda Y, Wataiki A, Watanabe S, Yamamura T, Yamanaka I, Yano R, Yasunishi A, Yokota T, Yoshida K, Yoshiki A, Yoshino M, Muramatsu M, and Hayashizaki Y

TITLE	RIKEN Mouse ESTs (Konno, H., et al.)
JOURNAL	Unpublished (2000)
COMMENT	Contact: Yoshinide Hayashizaki

genome Exploration Research Group, Life Science, Tsukuba Center,
genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center, Tsukuba, Japan
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: +81-298-36-9013
Fax: +81-298-36-9098

Email: genome-res@rtc.riken.go.jp,
URL: <http://genome.rtc.riken.go.jp/>

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Thermostabilization and thermoinactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)

Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303.

19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for

FEATURES location/qualifiers further details.

Source

```
/organism="Mus musculus"  
/db_xref="taxon:10090"  
/clone="C230016E14"  
/clone_1db="RIKEN full-length enriched, 0 day neonate  
cerebellum"
```

/lissue_type="cerebellum"
 /dev_stage="0 day neonate"
 /lab_host="DH10B"
 /note="Site_1: Sali; Site_2: BamHI: cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15' GAGGAGAGAGAGATCCACAGAGCTCTTTTTTTTTTTTTTTVN 3'. cDNA was prepared by using triethanolamine thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 479.0. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGAGAGAGATTCGACGTTAAATTAATATCCCCCCCCCC 3'. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLX I."

BASE COUNT	90 a	72 c	66 g	84 t
ORIGIN				
Query Match		45.9%	Score 29.4:	DB 129; Length 312;
Best Local Similarity		70.9%;	Pred. No. 3.5:	
Matches	39;	Conservative	0;	Mismatches 16; Indels 0; Gaps 0;
QY	5	cccaagtcgtctgaagcagcagcactagagctctctgcctgcctctctctcc	59	
Db	101	CCCACTGCTGTACTTACCTTAAGCTTACTACAGCTCCTCAGCTGGGAGGCTTCTTCTCC	47	
RESULT	8			
LOCUS	BEA15895/c			
DEFINITION		BEA15895	746 bp	mRNA
ACCESSION		MUG0002.B08R990520	ITEC	MUG Wheat Spikeleret Library Triticum aestivum
VERSION		BEA15895		
KEYWORDS		BEA15895.1	GI:9413741	
SOURCE				EST.
ORGANISM				bread wheat.
				Triticum aestivum
				Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
				Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.
REFERENCE				1 (bases 1 to 746)
AUTHORS				Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier
				, Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,
				Langridge, P., Izzo, G.R., Lin, J.J., McGuire, P., Ogihara, T.,
				Pechion, N., Quiset, C., Schuch, W., Selvaraj, G., Sharifou, M.,
				Sorrells, M., Warburton, M., and Wezel, G.
				International Triticale EST Cooperative (ITEC): Production of
				Expressed Sequence Tags for Species of the Triticaceae
				Unpublished (2000)
JOURNAL				Contact: Ogihara T
COMMENT				Kihara Institute for Biological Research, Yokohama City University
				Maoka-cho 641-12, Totsuka-ku, Yokohama 244-0813, JAPAN
				Tel: 81 45 820 1903
				Fax: 81 45 820 1901
				Email: ogihara@yokohama-cu.ac.jp
				International Triticale EST Cooperative (ITEC)
				http://wheat.pw.usda.gov/genome.
FEATURES				Location/Qualifiers
source				1..746
				/organism="Triticum aestivum"
				/cultivar="Norin 26"
				/db_xref="taxon:4565"
				/clone="MUG002.B08"
				/clone_lib="ITEC MUG Wheat Spikeleret Library"
				/tissue_type="young spikelets"
				/dev_stage="Peakes" scale 6-7"
				/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

BASE COUNT	ORIGIN	Short: M13 Reverse sequencing primer used. 1.2 Kbp average insert size."
BASE COUNT	175 a 236 c 208 g 125 t	2 others
Query Match	42.8% Score 27.4; DB 107; Length 746;	
Best Local Similarity	65.6% Pred. No. 19;	
Matches 40;	Conservative 0; Mismatches 21; Indels 0; Gaps 0;	
Oy	4 gccacagtgtcttagtgactaagcgcactagacattctcttcgtcgtcttccttcgcgc 63	
Db	194 gccacagccgcgcagccgccttccggccttgcgccttcttccgcgcgcgccttcacacgtc 135	
Oy	64 a 64	
Db	134 A 134	
RESULT 9		
LOCUS	A0933849 221 bp DNA GSS 21-DEC-1999	
DEFINITION	RPCT-23-289119.TV RPCT-23 Mus musculus genomic clone RPCT-23-289119	
ACCESSION	A0933849	
VERSION	A0933849	
KEYWORDS	' DNA sequence.	
SOURCE	A0933849.1 GI:6622863	
ORGANISM	GSS.	
	house mouse.	
	Mus musculus	
REFERENCE	Eukaryotic Metazoa: Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 221)	
	Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S., Akiret, B., Levins, M., McGann, S., Tesgeye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.	
TITLE	Mouse BAC End Sequences from Library RPCT-23	
JOURNAL	Unpublished (1999)	
COMMENT	Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: schaeftlgr.org Clones are derived from the mouse BAC library RPCT-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Reseach Genetics (Info@resgen.com). BAC end plate: http://www.ligr.org/ldb/bac_ends/mouse/bac_end_intro.html Plate: 289 row: I column: 19 Seq primer: T7 Class: BAC ends.	
FEATURES		
SOURCE	Location/Qualifiers	
	1..221	
	/organism="Mus musculus"	
	/strain="C57BL/6J"	
	/db_xref="taxon:10090"	
	/clone="RPCT-23-289119"	
	/clone_1lb="RPCT-23"	
	/sex="Female"	
	/lab_host="DH10B"	
	/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)." "	
BASE COUNT	47 a 47 c 46 g 76 t	5 others
ORIGIN		

C	22	36.4	15.0	14.6903	57	AC0686648	Homo sapi
C	23	35.8	14.8	15.4012	65	AC0709118	Homo sapi
C	24	35.8	14.8	16.2560	68	AL1358874	Homo sapi
C	25	35.8	14.8	17.0729	58	AC0692921	Homo sapi
C	26	35.8	14.8	17.9323	69	AL1351534	Homo sapi
C	27	35.8	14.8	19.6226	54	AC0320007	Homo sapi
C	28	35.8	14.8	22.1478	65	AC0795545	Homo sapi
C	29	35.6	14.7	19.6866	38	AC0107442	Homo sapi
C	30	35.4	14.6	85.620	49	AC021298	Homo sapi
C	31	35.4	14.6	93.540	8	AB042235	Homo sapi
C	32	35.4	14.6	126.334	66	AL133403	Homo sapi
C	33	35.4	14.6	128.241	8	AB041340	Homo sapi
C	34	35.4	14.6	181.302	47	AL1355532	Homo sapi
C	35	35.4	14.6	300.000	38	AP002530	Homo sapi
C	36	35.2	14.5	19.3227	38	AC010381	Homo sapi
C	37	35	14.5	15.4256	58	AC073295	Homo sapi
C	38	35	14.5	17.0356	11	AF259072	Homo sapi
C	39	35	14.5	201.833	84	CNS057C1	Homo sapi
C	40	34.8	14.4	16.5423	29	AC008111	Homo sapi
C	41	34.8	14.4	17.0216	19	AC012315	Homo sapi
C	42	34.8	14.4	17.3661	50	AC074107	Homo sapi
C	43	34.8	14.4	17.5789	69	AP000425	Homo sapi
C	44	34.6	14.3	11.1507	30	AC010001	Homo sapi
C	45	34.6	14.3	16.7213	58	AC069549	Homo sapi

ALIGNMENTS

RESULT	1						
LOCUS	AR037064						
DEFINITION	Sequence from patent US 5801016.					PAT	29-SEP-1999
ACCESSION	AR037064						
VERSION	AR037064.1						
KEYWORDS	GI:5954920						
SOURCE	Unknown.						
ORGANISM	Unknown.						
REFERENCE	Unclassified.						
AUTHORS	1 (bases 1 to 173)						
TITLE	Morioka,S. and Ueki,J.						
JOURNAL	DNA fragment, recombinant vector containing the same and method for						
FEATURES	expressing foreign genes using the same						
SOURCE	Patent: US 5801016-A 7 01-SEP-1998;						
	Location/Qualifiers						
	I..173						
	/organism="unknown"						
BASE COUNT	21 a	48 c	47 g		57 t		
ORIGIN							

Query Match	71.18;	Score 172;	DB 81;	length 173;
Best Local Similarity	100.00;	Pred. NO. 2.7e-39;		
Matches 172; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

[illegible]

VERSION	AR037061.1	GI:5954917		
KEYWORDS				
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 183)			
TITLE	Morioka,S. and Ueki,J.			
JOURNAL	DNA fragment, recombinant vector containing the same and method for			
FEATURES	expressing foreign genes using the same			
source	Patent: US 5801016-A 1 01-SEP-1998;			
	Location/Qualifiers			
	1..183			
	/organism="unknown"			
BASE COUNT	23 a	51 c	51 g	58 t
ORIGIN				

Query Match	71.1%	Score 172;	DB 81;	Length 183;
Best Local Similarity	100.0%	Pred. No. 2.7e-39;		
Matches 172; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	71	taagcccaagctgtgttagagccaagcaacaaagctcttgcgcagctgtctcttcag	130
Db	7	TAAACCCAGAGTGTCTTAGGCTAAAGCGACATGAGACTTGTGCTCGGTGCTTTCTCCG	66
QY	131	ctcagacatctgtctctgtctgtctgcctagaaccctactctgtctgcagctgcctcg	190
Db	67	CTCAGATCTGCTTTCCTTGGCTTCCCTTGAACCCCTACTCTGTGCTGCACAGTGTGCTG	126
QY	191	cttcagctctccctccccaagcttcagatctgtatctgtgtggtggtgggagccag	242
Db	127	CTTGCTCTCTCTTCCTCAAGTTCGATGTGATTGTGTGTGTGGGGGGGGCGCAG	178

LOCUS	AR005013	2799 bp	DNA	PAT	04-DEC-1998
DEFINITION	Sequence 5 from patent US 5747327.				
FEATURES					

RESULT	3				
LOCUS	AR005013				
DEFINITION	AR005013	2799 bp	DNA		PAT
ACCESSION	AR005013	Sequence 5 from patent US 5747327.			
VERSION	AR005013.1	GI:3965892			
KEYWORDS	.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 2799)				
AUTHORS	Ueki, J. and Morioaka, S.				
TITLE	Phospholipase D gene originated from plant				
JOURNAL	Patent: US 5747327-A 5 05-May-1998;				
FEATURES	Location/Qualifiers				
source	1..2799				
	/organism="unknown"				
BASE COUNT	692 a	709 c	609 g		789 t
ORIGIN					

Query Match	71.1%	Score 172;	DB 81;	Length 2799;
Best Local Similarity	100.0%	Pred. NO. 3e-39;		
Matches 172; Conservative	0;	Mismatches	0;	Gaps 0;

QY	130
71	taacccagatgagctcttaggcgaagcgacataaagctctctgtgcgttcgtctctctccg
1667	taaacccagatgagctcttaggcgaagcgacataaagctctctgtgcgttcgtctctccg
QY	190
131	ctcagatctgctgtcgtctgtctgcgtacagaccctactctgtctgcgtgagtgctg
1727	ctcagatctgctgtcgtctgtctgcgtacagaccctactctgtctgcgtgagtgctg
Db	ctcagatctgctgtcgtctgtctgcgtacagaccctactctgtctgcgtgagtgctg
QY	242
191	cttcgctcttcctctccccaagtcgcagatctgattgtgtgtgtagggggcgacg
1787	cttcgctcttcctctccccaagtcgcagatctgattgtgtgtgtagggggcgacg
Db	cttcgctcttcctctccccaagtcgcagatctgattgtgtgtgtagggggcgacg
RESULT	4

AR037063 AR037063 2799 bp DNA PAT 29-SEP-1999
LOCUS
DEFINITION Sequence 4 from patent US 5801016.
ACCESSION AR037063
VERSION AR037063.1 GI:5954919
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
BASE COUNT 692 a 709 c 609 g 789 t
ORIGIN

Query Match 71.1%; Score 172; DB 81; Length 2799;
Best Local Similarity 100.0%; Pred. No. 3e-39;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 71 taagccagtgctgctgaagcgaacgacatgctgctgctgctgctcgcg 130
|||||
Db 1667 TAAGCCAGTGTGCTTAGGCTAAGCGCAGCTGCTTCTGCTTCTTCGC 1726
|||||

Oy 131 ctccagatcgtgctgctgctgctgctgaacacctactctgtgctgagatgctg 190
|||||
Db 1727 CTCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1786
|||||

Oy 191 ctgctgctctctctcaagtcgctgctgctgctgctgctgctgctgctg 242
|||||
Db 1787 CTGCTCTCTCTCTCAAGTTCGATTCGATTCGATTCGATTCGATTCG 1838
|||||

RESULT 5
AB001920
LOCUS AB001920 5871 bp DNA PLN 14-APR-2000
DEFINITION Oryza sativa DNA for phospholipase D, complete cds.
ACCESSION AB001920
VERSION AB001920.1 GI:1902902
KEYWORDS phospholipase D.
SOURCE Oryza sativa (cultivar:koshihikari) leaf DNA.
ORGANISM Oryza sativa
Enkaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE
AUTHORS Ueki,J., Morioka,S., Komari,T. and Kumashtiro,T.
TITLE 1 (bases 1522 to 1665; 1839 to 1983; 2524 to 4420)
JOURNAL Plant Cell Physiol. 36 (5), 903-914 (1995)
MEDLINE 96012933
JOURNAL MEDLINE 2 (bases 1 to 5871)
REFERENCE 96012933
AUTHORS Morioka,S., Ueki,J. and Komari,T.
TITLE Characterization of two distinctive genomic clones (Accession Nos.
AB001919 and AB001920) for phospholipase D from rice (PGR97-076)
JOURNAL Plant Physiol. 114, 396 (1997)
REFERENCE 3 (bases 1 to 5871)
AUTHORS Ueki,J.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1995) to the DDBJ/EMBL/Genbank databases. Jun
Ueki, Japan Tobacco Inc., Plant Breeding and Genetics Research Lab;
700 Higashibara, Iwata, Shizuoka 438-0802, Japan
(E-mail:Jun.ueki@pbgrl.jti.co.jp, Tel:81-538-32-7111,
Fax:81-538-32-8700)
FEATURES
SOURCE location/Qualifiers
1..5871
/organism="Oryza sativa"
/cultivar="Koshihikari"

/db_xref="taxon:4530"
/tissue_type="leaf"
join(1522..1665,1839..1983,2524..4420,4899..5702)
/citation=[1]
1522..1665
/number=1
/number=1
1666..1838
/number=1
1839..1983
/number=2
join(1876..1983,2524..4420,4899..5332)
/EC_number="3.1.4.4"
/citation=[1]
/codon_start=1
/product="phospholipase D"
/protein_id="BAA19467.1"
/db_xref="GI:1902903"
/translation="MAOMLLGLTLHATIFEPAASLSINPRAAGSAPKFIKRVGIEDT
VGWKGATKVVSTIDLEKARVGRMTNEPIMPWRESFHIYCAHMASNVITVKID
NPIGATNIGRAYLPVDELNGEIDRLDIDNNRESVGEISKIHKLYQDVASKRNM
ARGVSRTPKVPYPTFEFSOROGCKVTLYDAHADPNEIPKIPLADGKNYPHRCMDI
FDAISNAOHLITITGWSVYETITLYVRDNRNPKGCGVTGICELLKKASBEGRVLMLYM
DDRTSVGLLRDGLMATHDEETERYFHGSVNCVLCPRNPDGSGTIVQDLSITMFT
HQRITVVVDHLEPNQGSQORIVSFVGGDLCDGXYDTQYHSLERTLDSTHDEHQP
FATASIKKGGPREPMDIHSLBSPIMADVLVNEPQWRKQGGKDLQIRDSDTII
PSPVMPPEDEETNNVOLFRSIDGAFGPPDPPEAKAGLVSGKDOIIDRSIDQAY
IHAIIRAKNFITYENOVFICSSVAKRPGIKPEDIGLHLIPKELALKVSKLEAGR
FYIVVYVPMPEGVPEGSQAILDMQRRTENMYTDTIALQAKGTEANPKDTLTFE
CLGKREKQAGEFQPEQPADVDYSAQDAARRMTYVHKRMIVDEYIITISANT
QKSMDEARDESIAMGQYPLHATROAPARHIGFRAVLAHEHGLMDVDFORPSE
CVOKVNRIAEKYMDMYSDDLODLPGLHLSYPLIGVASDGVVELPMEYFPDTRARV
LGASDMPIPLTS"
1984..2523
/number=2
2524..4420
/number=3
4421..4898
/number=3
4899..5702
/number=4
BASE COUNT 1488 a 1363 c 1355 g 1665 t
ORIGIN

Query Match 71.1%; Score 172; DB 6; Length 5871;
Best Local Similarity 100.0%; Pred. No. 3e-39;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 71 taagccagtgctgctgaagcgaacgacatgctgctgctgctgctcgcg 130
|||||
Db 1667 TAAGCCAGTGTGCTTAGGCTAAGCGCAGCTGCTTCTGCTTCTTCGC 1726
|||||

Oy 131 ctccagatcgtgctgctgctgctgctgaacacctactctgtgctgagatgctg 190
|||||
Db 1727 CTCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1786
|||||

Oy 191 ctgctgctctctctcaagtcgctgctgctgctgctgctgctgctgctg 242
|||||
Db 1787 CTGCTCTCTCTCTCAAGTTCGATTCGATTCGATTCGATTCGATTCG 1838
|||||

RESULT 6
I66494
LOCUS I66494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Unknown.
Unclassified.
1 (bases 1 to 7218)
Dorner,F., Schefflenger,F. and Falkner,F. Gunter.


```
* 44143 44242: gap of 100 bp
* 44243 45375: contig of 1133 bp in length
* 45376 45475: gap of 100 bp
* 45476 48030: contig of 2555 bp in length
* 48031 48130: gap of 100 bp
* 48131 50762: contig of 2632 bp in length
* 50763 50862: gap of 100 bp
* 50863 54865: contig of 4003 bp in length
* 54866 54965: gap of 100 bp
* 54966 58786: contig of 3821 bp in length
* 58787 58886: gap of 100 bp
* 58887 62073: contig of 3193 bp in length
* 62080 62179: gap of 100 bp
* 62180 66768: contig of 4589 bp in length
* 66769 66868: gap of 100 bp
* 66869 70876: contig of 4008 bp in length
* 70877 70976: gap of 100 bp
* 70977 75996: contig of 5020 bp in length
* 75997 76096: gap of 100 bp
* 76097 80273: contig of 4177 bp in length
* 80274 80373: gap of 100 bp
* 80374 85679: contig of 5306 bp in length
* 85680 85779: gap of 100 bp
* 85780 91717: contig of 5938 bp in length
* 91718 91817: gap of 100 bp
* 91818 97046: contig of 5229 bp in length
* 97047 97146: gap of 100 bp
* 97147 101127: contig of 3981 bp in length
* 101128 101227: gap of 100 bp
* 101228 106788: contig of 5561 bp in length
* 106789 106888: gap of 100 bp
* 106889 111554: contig of 4666 bp in length
* 111555 111654: gap of 100 bp
* 111655 118598: contig of 6944 bp in length
* 118599 118698: gap of 100 bp
* 118699 124207: contig of 5509 bp in length
* 124208 124307: gap of 100 bp
* 124308 130324: contig of 6017 bp in length
* 130325 130424: gap of 100 bp
* 130425 138528: contig of 8104 bp in length
* 138529 138628: gap of 100 bp
* 138629 146991: contig of 8363 bp in length
* 146992 147091: gap of 100 bp
* 147092 153531: contig of 6440 bp in length
* 153532 153631: gap of 100 bp
* 153632 162631: contig of 9000 bp in length
* 162632 162731: gap of 100 bp
* 162732 174676: contig of 11945 bp in length
* 174677 174776: gap of 100 bp
* 174777 186910: contig of 12134 bp in length
* 186911 187010: gap of 100 bp
* 187011 201251: contig of 14241 bp in length.
*
FEATURES
SOURCE
1..201251
Location/Qualifiers
1..201251
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/map="11"
/clone="RP23-136C7"
/clone_id="RP23 Female Mouse BAC"
1..1487
misc_feature
/note="assembly_fragment"
1588..2793
misc_feature
/note="assembly_fragment"
2894..4247
misc_feature
/note="assembly_fragment"
4348..5849
misc_feature
/note="assembly_fragment"
5950..7248
misc_feature
/note="assembly_fragment"
7349..8876
misc_feature
/note="assembly_fragment"
8977..10439
```

```
misc_feature
/note="assembly_fragment"
10540..12080
misc_feature
/note="assembly_fragment"
12181..14702
misc_feature
/note="assembly_fragment"
14803..16733
misc_feature
/note="assembly_fragment"
16834..18755
misc_feature
/note="assembly_fragment"
18856..20786
misc_feature
/note="assembly_fragment"
20887..23125
/note="assembly_fragment"

Query Match 18.2% Score 44; DB 52; Length 201251;
Best Local Similarity 51.0%; Pred. No. 0.023;
Matches 104; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 36 ctcttgctgctgcttcttcctcgcctcagatccctaaagccagtgctagagc 95
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 42170 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 42111

QY 96 gcactagagcttctgctcgctgcttcttcctcgcctcagatcgtgctgctt 155
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 42110 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 42051

QY 156 cgcctagaccctactcctgctgctcagagtgctgctgcttccttcctcctcagctga 215
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 42050 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 41991

QY 216 tctgattgtgtgtggtggggggcg 239
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41990 TGTGTGTGTGTGTGTGTGTGTG 41967

RESULT 8
AC026682 204806 bp DNA HTG 06-OCT-2000
LOCUS
DEFINITION
AC026682.9 GI:10645292
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 204806)
Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okunou,G., Carlack,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
Gosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Louised,H., Lozano,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogura,M., Parish,B.,
Perez,L., Reltter,D., Say,J., Shen,X., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstein,G.,
Worley,K. and Gibbs,R.
Direct Submission
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
```



```
* 1279 1431: contig of 153 bp in length
* 1432 1531: gap of 100 bp
* 1532 2672: contig of 1141 bp in length
* 2673 2772: gap of 100 bp
* 2773 3853: contig of 1081 bp in length
* 3854 3953: gap of 100 bp
* 3954 5494: contig of 1541 bp in length
* 5495 5594: gap of 100 bp
* 5595 6808: contig of 1214 bp in length
* 6809 6908: gap of 100 bp
* 6909 8821: contig of 1913 bp in length
* 8822 8921: gap of 100 bp
* 8922 10227: contig of 1306 bp in length
* 10228 10327: gap of 100 bp
* 10328 12517: contig of 2090 bp in length
* 12518 14413: contig of 1896 bp in length
* 14414 14513: gap of 100 bp
* 14514 16780: contig of 2267 bp in length
* 16781 16880: gap of 100 bp
* 16881 18989: contig of 2109 bp in length
* 18990 19089: gap of 100 bp
* 19090 21339: contig of 2250 bp in length
* 21340 21439: gap of 100 bp
* 21440 24496: contig of 3057 bp in length
* 24497 24596: gap of 100 bp
* 24597 27488: contig of 2892 bp in length
* 27489 27588: gap of 100 bp
* 27589 30111: contig of 2523 bp in length
* 30112 30211: gap of 100 bp
* 30212 33587: contig of 3376 bp in length
* 33588 33687: gap of 100 bp
* 33688 37578: contig of 3891 bp in length
* 37579 37678: gap of 100 bp
* 37679 42795: contig of 5117 bp in length
* 42796 42895: gap of 100 bp
* 42896 48285: contig of 5390 bp in length
* 48286 48385: gap of 100 bp
* 48386 52171: contig of 3786 bp in length
* 52172 52271: gap of 100 bp
* 52272 60414: contig of 8143 bp in length
* 60415 60514: gap of 100 bp
* 60515 68386: contig of 7872 bp in length
* 68387 68486: gap of 100 bp
* 68487 75220: contig of 6734 bp in length
* 75221 75320: gap of 100 bp
* 75321 82520: contig of 7200 bp in length
* 82521 82620: gap of 100 bp
* 82621 92012: contig of 9392 bp in length
* 92013 92112: gap of 100 bp
* 92113 100585: contig of 8473 bp in length
* 100586 100685: gap of 100 bp
* 100686 111797: contig of 11112 bp in length
* 111798 111897: gap of 100 bp
* 111898 120279: contig of 8382 bp in length
* 120280 120379: gap of 100 bp
* 120380 132436: contig of 12057 bp in length
* 132437 132536: gap of 100 bp
* 132537 144076: contig of 11540 bp in length
* 144077 144176: gap of 100 bp
* 144177 160302: contig of 16126 bp in length
* 160303 160402: gap of 100 bp
* 160403 175068: contig of 14666 bp in length
* 175069 175168: gap of 100 bp
* 175169 192239: contig of 17071 bp in length
* 192240 192339: gap of 100 bp
* 192340 207757: contig of 15418 bp in length.
```

FEATURES
Source

```
1. 207757
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
```

```
/clone="RP11-618p13"
/clone_id="RP11 Human Male BAC"
1. 1178
/feature="assembly-fragment"
1279. 1431
/feature="assembly-fragment"
vector_end:sp6
vector_side:right"
1532. 2672
/feature="assembly-fragment"
2773. 3853
/feature="assembly-fragment"
3954. 5494
/feature="assembly-fragment"
5595. 6808
/feature="assembly-fragment"
6909. 8821
/feature="assembly-fragment"
8922. 10227
/feature="assembly-fragment"
10328. 12417
/feature="assembly-fragment"
12518. 14413
/feature="assembly-fragment"
14514. 16780
/feature="assembly-fragment"
16881. 18989
/feature="assembly-fragment"
19090. 21339
/feature="assembly-fragment"
21440. 24496
/feature="assembly-fragment"
24597. 27488
/feature="assembly-fragment"
27589. 30111
/feature="assembly-fragment"
30212. 33587
/feature="assembly-fragment"
33688. 37578
/feature="assembly-fragment"
37679. 42795
/feature="assembly-fragment"
42896. 48285
/feature="assembly-fragment"
48386. 52171
/feature="assembly-fragment"
52272. 60414
/feature="assembly-fragment"
60515. 68386
/feature="assembly-fragment"
68487. 75220
/feature="assembly-fragment"
75321. 82520
/feature="assembly-fragment"
```

Query Match 17.9% Score 43.2; DB 51; Length 207757;
Best Local Similarity 51.0%; Pred. No. 0.039; Mismatches 98; Indels 0; Gaps 0;

```
QY 36 cttctgctgctgcttcttctcgcgtcagatccaaagccagtgctgtagtaagc 95
||||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 143877 CTTCTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 143936

QY 96 gcaactagagctcctgctgcttcttctcgcgcagatcgtctgcttgcctt 155
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 143937 CTTCTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 143996

QY 156 gcctgaaccctacacctgctgctgctgctgctgcttcttcttcttcttcttctga 215
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 143997 CTTCTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 144056

QY 216 tctgattgtgtgtgtgtggtg 235
||| ||| ||||| ||| |||
```



```
----- Summary Statistics -----
Assembly program: XGAP; version 4.5
Sequencing vector: Plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 169222 bases at least Q40
Consensus quality: 170056 bases at least Q30
Consensus quality: 170433 bases at least Q20
Insert size: 170741; sum-of-contigs
Insert size: 171184; 25.5% error; agarose-fp
Quality coverage: 5.17x in Q20 bases; sum-of-contigs
Quality coverage: 5.57x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3832: contig of 3832 bp in length
* 3833 3932: gap of 100 bp
* 3933 165252: contig of 161320 bp in length
* 165253 165352: gap of 100 bp
* 165353 167736: contig of 2384 bp in length
* 167737 167836: gap of 100 bp
* 167837 171041: contig of 3205 bp in length.
*
* Location/Qualifiers
*   1..171041
*     /organism="Homo sapiens"
*     /db_xref="taxon:9606"
*     /chromosome="13"
*     /clone="RP11-185C18"
*     /clone_1fb="RPC1-11.1"
*     1..3832
*       /note="assembly_fragment:00817"
*       fragment_chain:1
*         clone_end:SP6
*         vector_side:left
*         3933..165252
*           /note="assembly_fragment:02691"
*           fragment_chain:1
*             165353..167736
*             /note="assembly_fragment:02301"
*             167837..171041
*             /note="assembly_fragment:02333"
*
* BASE COUNT  5068 a 34334 c 35684 g 50052 t 303 others
* ORIGIN
*
* Query Match 16.3%; Score 39.4; DB 66; Length 171041;
* Best Local Similarity 51.4%; Pred. No. 0.48;
* Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
*
* Oy 1 gtaagccagtgctgtaagcgacactagactctgtcgtctctcc 60
*   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
* Db 19991 GTAGAGTAATGAGAGCTAGTAACTGCAAGTGGGATACAGATATTTCAACTGCCA 20050
*
* Oy 61 gctcagatccctaaagccagtgctgtaagcgacactagactctgtcgtctgc 120
*   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
* Db 20051 GAGAAAATACGACAGCAACATGGCTTGCAATATATCTTAAAGACGCTTACACTCTTTC 20110
*
* Oy 121 tctctcgcgtcagatctgtctgtctgtcgtctgtagaacctactctgtctg 177
*   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
* Db 20111 TACCTCTCTTGCAATTTTGCGTCTCTTTCTTCTTCTTGACACCTGTATTATCT 20167
*
* RESULT 12
* AC079744 AC079744 135697 bp DNA HTG 10-SEP-2000
* LOCUS Homo sapiens chromosome UNK clone CTD-2326K17, *** SEQUENCING IN
* DEFINITION PROGRESS ***, 31 unordered pieces.
* ACCESSION AC079744
```

```
VERSION AC079744.1 GI:10047923
KEYWORDS HTG; HTGS_PHASEL.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 135697)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 135697)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_MS2326K17
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big Dye; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 116218 bases at least Q40
Consensus quality: 123738 bases at least Q30
Consensus quality: 126348 bases at least Q20
Insert size: 104000; agarose-fp
Insert size: 132052; sum-of-contigs
Quality coverage: 4.24 in Q20 bases; agarose-fp
Quality coverage: 3.44 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 645: contig of 645 bp in length
* 646 745: gap of unknown length
* 746 1955: contig of 1210 bp in length
* 1956 2055: gap of unknown length
* 2056 3319: contig of 1264 bp in length
* 3320 3419: gap of unknown length
* 3420 4702: contig of 1283 bp in length
* 4703 4802: gap of unknown length
* 4803 6106: contig of 1304 bp in length
* 6107 6206: gap of unknown length
* 6207 7592: contig of 1386 bp in length
* 7593 7692: gap of unknown length
* 7693 9298: contig of 1606 bp in length
* 9299 9398: gap of unknown length
* 9399 10800: contig of 1402 bp in length
* 10801 10900: gap of unknown length
* 10901 12209: contig of 1309 bp in length
* 12210 12309: gap of unknown length
* 12310 13776: contig of 1467 bp in length
* 13777 13876: gap of unknown length
* 13877 15031: contig of 1155 bp in length
* 15032 15131: gap of unknown length
* 15132 16621: contig of 1490 bp in length
* 16622 16721: gap of unknown length
* 16722 18441: contig of 1720 bp in length
* 18442 18541: gap of unknown length
* 18542 20385: contig of 1844 bp in length
* 20386 20485: gap of unknown length
```


REFERENCE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 144239)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
Summary Statistics
Center project name: H_RG333f24

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7> or <mailto:egreen@nhgri.nih.gov>

SOURCE INFORMATION:
Clone CTA-333f24 is from a release of the human BAC library CTB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBel0BAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is CTA-242D1, 200 bp overlap. The actual start of this clone is at base position 195 of CTA-333f24; actual end is at 145016 of CTA-333f24. This clone is part of an unanchored island, orientation is unknown.

The clone CTA-333f24 contains a transposable element, from bacterial DNA, starting at position 33810 and ending at 34586.

FEATURES

SOURCE
Location/Qualifiers
1..144239
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="CTA-333f24"
/clone_1bp="CTB-HS-A"
/map="7q11.2-q21"
531..829
/rpt_family="Alu"
1072..1484
/rpt_family="MaLR"
1576..2103
/rpt_family="Retroviral"
3406..3596
/rpt_family="MIR"
4315..4481
/rpt_family="L2"
5683..5741
/rpt_family="L2"
5844..6210
repeat_region

/rpt_family="Retroviral"
7289..7376
/rpt_family="Achobo"
7384..7851
/rpt_family="MaLR"
7852..8374
/rpt_family="Achobo"
8373..8593
/rpt_family="Achobo"
8594..8727
/rpt_family="Alu"
8728..8898
/rpt_family="Achobo"
8988..9308
/rpt_family="Retroviral"
9543..9638
/rpt_family="MIR"
9638..9805
/rpt_family="Alu"
11635..11813
/rpt_family="MIR"
12320..12534
/rpt_family="MIR"
12751..13078
/rpt_family="MER2_type"
15243..15393
/rpt_family="L1"
16127..16486
/rpt_family="MaLR"
16487..18049
/rpt_family="MaLR"
18052..18390
/rpt_family="MaLR"
21329..22032
/rpt_family="L2"
28704..28821
/rpt_family="L1"
30594..30661
/rpt_family="Retroviral"
32037..32156
/rpt_family="L2"
33691..33809
/rpt_family="L1"
33828..34184
/rpt_family="MaLR"
34247..34513
/rpt_family="Alu"
34516..35404
/rpt_family="L1"
36874..37266
/rpt_family="MaLR"
37721..38010
/rpt_family="Alu"
39885..40072
/rpt_family="MIR"
40109..40407
/rpt_family="Alu"
41954..42551
/rpt_family="L1"
42639..42713
/rpt_family="L1"
43553..46528
/rpt_family="L1"
47407..47901
/rpt_family="MER21_g"
47929..48048
/rpt_family="L2"
50475..50560
/rpt_family="L2"
52156..52208
/rpt_family="L2"
52380..53628
/rpt_family="L1"
repeat_region

```
repeatL_region      53605. .53682  
/rpl_family="L1"  
repeatL_region      53691. .53959  
/rpl_family="Alu"  
repeatL_region      54014. .54574  
/rpl_family="L1"  
repeatL_region      54598. .54857  
/rpl_family="L2"  
repeatL_region      55086. .55202  
/rpl_family="L1"  
repeatL_region      55507. .55758  
/rpl_family="Alu"  
repeatL_region      56035. .56438  
/rpl_family="MER1_type"  
repeatL_region      57047. .57517  
/rpl_family="MER1_type"  
repeatL_region      57741. .58049  
/rpl_family="Alu"  
repeatL_region      58651. .59146  
/rpl_family="MER1"  
repeatL_region      61023. .61181  
/rpl_family="MER1_type"  
repeatL_region      61994. .62142  
/rpl_family="Alu"  
repeatL_region      62212. .62510  
/rpl_family="Alu"  
repeatL_region      63961. .64059  
/rpl_family="MER1_type"  
repeatL_region      64048. .64214  
/rpl_family="MER1_type"  
repeatL_region      65304. .65390  
/rpl_family="L2"
```

Query Match	15.8%;	Score 38.2;	DB 8;	Length 144239;
Best Local Similarity	54.7%;	Pred. No. 1;		
Matches 76;	Conservative	0;	Mismatches 63;	Indels 0;
				Gaps 0;

[illegible]

RESULT	14
LOCUS	AC069368
DEFINITION	AC069368 116534 bp DNA HTG 17-JUN-2000
ACCESSION	HOMO_SAPIENS_chromosome_15_clone_CTD-2017F17 map 15, WORKING DRAFT
VERSION	AC069368.2 GI:8570510
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 116534)
TITLE	Birtten,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL	Homo sapiens chromosome 15, clone CTD-2017F17
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 116534)
AUTHORS	Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

1 (bases 1 to 116534)
Birren, B., Linton, L., Nussbaum, C. and Lander, E.
Homo sapiens chromosome 15, clone CTD-2017F17
Unpublished
2 (bases 1 to 116534)
Birren, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barina, N., Bastien, V., Beda, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burnett, G.,
Campomiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearlilano, K., Dewar, K., Diaz, J. S.,
Dodge, S., Dominko, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,

TITLE
JOURNAL
COMMENT

TITLE

JOURNAL

COMMENT

```

Galagan,J.,Gardyna,S.,Ginde,S.,Goyette,M.,Graham,L.,
Grand-Pierre,N.,Grant,G.,Hagos,B.,Heaford,A.,Horton,L.,
Howard,J.C.,Iliev,I.,Johnson,R.,Jones,C.,Kann,L.,Karlas,A.,
Klein,J.,Lacroque,K.,Lamazez,R.,Landers,T.,Lehoczky,J.,
Levine,R.,Lieu,C.,Liu,G.,Locke,K.,Macdonald,P.,Margis,N.,
McCarthy,J.M.,McEwan,P.,McCur,A.,McKenra,C.,McPheeters,R.,
Meldrum,J.,Meneus,L.,Mihova,T.,Miranda,C.,Mlenga,V.,Morrow,J.,
Murphy,T.,Naylor,J.,Norman,C.H.,O'Connor,T.,O'Donnell,P.,
O'Neill,D.,Ollivier,T.M.,Olliver,J.,Peterson,K.,Pierre,N.,
Pisani,C.,Pollara,V.,Raymond,C.,Riley,R.,Rogov,P.,Rothman,D.,
Roy,A.,Santos,R.,Schauer,S.,Severy,P.,Spencer,B.,
Stange-Thomann,N.,Stojanovic,N.,Subramanian,A.,Talamas,J.,
Tessilev,S.,Theodore,J.,Tirrell,A.,Travers,M.,Trigillo,J.,
Vassiliev,H.,Viel,R.,Vo,A.,Wilson,B.,Wu,X.,Wyman,D.,Ye,W.J.,
Young,G.,Zainoun,J.,Zimmer,A. and Zody,M.
Direct Submission
Submitted (28-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 17, 2000 this sequence version replaced gi:8093855.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: LB893
Center clone name: 2017.F.17
-----
Summary Statistics
Sequencing vector: M13: M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 101291 bases at least Q40
Consensus quality: 108978 bases at least Q30
Consensus quality: 111706 bases at least Q20
Insert size: 133000; agarose-fp
Insert size: 113534; sum-of-contigs
Quality coverage: 2.8 in Q20 bases; agarose-fp
Quality coverage: 3.2 in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently
consists of 31 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
*
1
1030: contig of 1030 bp in length
*
1031 1130: gap of 100 bp
1131 2382: contig of 1252 bp in length
*
2383 2482: gap of 100 bp
2483 4018: contig of 1536 bp in length
*
4019 4118: gap of 100 bp
4119 5423: contig of 1305 bp in length
*
5424 5523: gap of 100 bp
5524 7327: contig of 1804 bp in length
*
7328 7427: gap of 100 bp
7428 8620: contig of 1193 bp in length
*
8621 8720: gap of 100 bp
8721 10395: contig of 1675 bp in length
*
10396 10495: gap of 100 bp
10496 12125: contig of 1630 bp in length
*
12126 12225: gap of 100 bp
12226 14121: contig of 1896 bp in length
*
14122 14221: gap of 100 bp
14222 16381: contig of 2160 bp in length
*
16382 16481: gap of 100 bp
16482 19335: contig of 2854 bp in length
*
19336 19435: gap of 100 bp
19436 21801: contig of 2366 bp in length

```

[illegible]

URES	source	Location/Qualifiers
misc_feature	1..116534	/organism="Homo sapiens"
misc_feature	/db_xref="taxon:9606"	/chromosome="15"
misc_feature	/map="15"	/clone="CPD-2017F17"
misc_feature	/clone_11b="CITD Human BAC"	1..1030
misc_feature	/note="assembly_fragment"	1131..2362
misc_feature	/note="assembly_fragment"	2483..4018
misc_feature	/note="assembly_fragment"	4119..5423
misc_feature	/note="assembly_fragment"	5524..7327
misc_feature	/note="assembly_fragment"	7428..8620
misc_feature	/note="assembly_fragment"	8721..10395
misc_feature	/note="assembly_fragment"	10496..12125
misc_feature	/note="assembly_fragment"	12226..14121
misc_feature	/note="assembly_fragment"	14222..16381
misc_feature	/note="assembly_fragment"	16482..19335
misc_feature	/note="assembly_fragment"	19436..21801
misc_feature	/note="assembly_fragment"	21902..23223
misc_feature	/note="assembly_fragment"	23324..26312
misc_feature	/note="assembly_fragment"	26312..28312
misc_feature	/note="assembly_fragment"	28312..30312
misc_feature	/note="assembly_fragment"	30312..32312
misc_feature	/note="assembly_fragment"	32312..34312
misc_feature	/note="assembly_fragment"	34312..36312
misc_feature	/note="assembly_fragment"	36312..38312
misc_feature	/note="assembly_fragment"	38312..40312
misc_feature	/note="assembly_fragment"	40312..42312
misc_feature	/note="assembly_fragment"	42312..44312
misc_feature	/note="assembly_fragment"	44312..46312
misc_feature	/note="assembly_fragment"	46312..48312
misc_feature	/note="assembly_fragment"	48312..50312
misc_feature	/note="assembly_fragment"	50312..52312
misc_feature	/note="assembly_fragment"	52312..54312
misc_feature	/note="assembly_fragment"	54312..56312
misc_feature	/note="assembly_fragment"	56312..58312
misc_feature	/note="assembly_fragment"	58312..60312
misc_feature	/note="assembly_fragment"	60312..62312
misc_feature	/note="assembly_fragment"	62312..64312
misc_feature	/note="assembly_fragment"	64312..66312
misc_feature	/note="assembly_fragment"	66312..68312
misc_feature	/note="assembly_fragment"	68312..70312
misc_feature	/note="assembly_fragment"	70312..72312
misc_feature	/note="assembly_fragment"	72312..74312
misc_feature	/note="assembly_fragment"	74312..76312
misc_feature	/note="assembly_fragment"	76312..78312
misc_feature	/note="assembly_fragment"	78312..80312
misc_feature	/note="assembly_fragment"	80312..82312
misc_feature	/note="assembly_fragment"	82312..84312
misc_feature	/note="assembly_fragment"	84312..86312
misc_feature	/note="assembly_fragment"	86312..88312
misc_feature	/note="assembly_fragment"	88312..90312
misc_feature	/note="assembly_fragment"	90312..92312
misc_feature	/note="assembly_fragment"	92312..94312
misc_feature	/note="assembly_fragment"	94312..96312
misc_feature	/note="assembly_fragment"	96312..98312
misc_feature	/note="assembly_fragment"	98312..100312
misc_feature	/note="assembly_fragment"	100312..102312
misc_feature	/note="assembly_fragment"	102312..104312
misc_feature	/note="assembly_fragment"	104312..106312
misc_feature	/note="assembly_fragment"	106312..108312
misc_feature	/note="assembly_fragment"	108312..110312
misc_feature	/note="assembly_fragment"	110312..112312
misc_feature	/note="assembly_fragment"	112312..114312
misc_feature	/note="assembly_fragment"	114312..116312
misc_feature	/note="assembly_fragment"	116312..118312
misc_feature	/note="assembly_fragment"	118312..120312
misc_feature	/note="assembly_fragment"	120312..122312
misc_feature	/note="assembly_fragment"	122312..124312
misc_feature	/note="assembly_fragment"	124312..126312
misc_feature	/note="assembly_fragment"	126312..128312
misc_feature	/note="assembly_fragment"	128312..130312
misc_feature	/note="assembly_fragment"	130312..132312
misc_feature	/note="assembly_fragment"	132312..134312
misc_feature	/note="assembly_fragment"	134312..136312
misc_feature	/note="assembly_fragment"	136312..138312
misc_feature	/note="assembly_fragment"	138312..140312
misc_feature	/note="assembly_fragment"	140312..142312
misc_feature	/note="assembly_fragment"	142312..144312
misc_feature	/note="assembly_fragment"	144312..146312
misc_feature	/note="assembly_fragment"	146312..148312
misc_feature	/note="assembly_fragment"	148312..150312
misc_feature	/note="assembly_fragment"	150312..152312
misc_feature	/note="assembly_fragment"	152312..154312
misc_feature	/note="assembly_fragment"	154312..156312
misc_feature	/note="assembly_fragment"	156312..158312
misc_feature	/note="assembly_fragment"	158312..160312
misc_feature	/note="assembly_fragment"	160312..162312
misc_feature	/note="assembly_fragment"	162312..164312
misc_feature	/note="assembly_fragment"	164312..166312
misc_feature	/note="assembly_fragment"	166312..168312
misc_feature	/note="assembly_fragment"	168312..170312
misc_feature	/note="assembly_fragment"	170312..172312
misc_feature	/note="assembly_fragment"	172312..174312
misc_feature	/note="assembly_fragment"	174312..176312
misc_feature	/note="assembly_fragment"	176312..178312
misc_feature	/note="assembly_fragment"	178312..180312
misc_feature	/note="assembly_fragment"	180312..182312
misc_feature	/note="assembly_fragment"	182312..184312
misc_feature	/note="assembly_fragment"	184312..186312
misc_feature	/note="assembly_fragment"	186312..188312
misc_feature	/note="assembly_fragment"	188312..190312
misc_feature	/note="assembly_fragment"	190312..192312
misc_feature	/note="assembly_fragment"	192312..194312
misc_feature	/note="assembly_fragment"	194312..196312
misc_feature	/note="assembly_fragment"	196312..198312
misc_feature	/note="assembly_fragment"	198312..200312
misc		

```
misc_feature /note="assembly_fragment" 26413. .28720
misc_feature /note="assembly_fragment" 28821. .31342
misc_feature /note="assembly_fragment" 31443. .34159
misc_feature /note="assembly_fragment" 34260. .35873
misc_feature /note="assembly_fragment" 35974. .38802
misc_feature /note="assembly_fragment" 38903. .41723
misc_feature /note="assembly_fragment" 41824. .45044
misc_feature /note="assembly_fragment" 45145. .48071
misc_feature /note="assembly_fragment" 48172. .52438
misc_feature /note="assembly_fragment" 52539. .55979
misc_feature /note="assembly_fragment" 56080. .60454
misc_feature /note="assembly_fragment" 60555. .65352
misc_feature /note="assembly_fragment" 65453. .69953
misc_feature /note="assembly_fragment" 70054. .77023
misc_feature /note="assembly_fragment" 77124. .86478
misc_feature /note="assembly_fragment" 86579. .98029
misc_feature /note="assembly_fragment"
```

Query Match	15.7%	Score 38	DB 58	Length 116534
Best Local Similarity	49.0%	Pred. No. 1.2		
Matches 101	Conservative	0	Mismatches 105	Indels 0
				Gaps
Oy	21	ctaaagcagactagagctctctgctgcgtctctctctcgcgcacagatccctaaagccagct 80		
Db	115546	CTTAGTTCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTC 115605		
Oy	81	gtgcttaggctaaagcagactagagactctctgctgcgtctctctctctcgcgcacagatcg 140		
Db	115606	TCTCCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 115665		
Oy	141	cttgctctgctgctgcctgcgtctagaaacctactctctgctgcgagtgctgcgtctctcttc 200		
Db	115666	CTTCTCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 115725		
Oy	201	cttcctcaagctgcagctcgtatgctgt 226		
Db	115726	CCTCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 115751		
RESULT 15				
AC009602/c				
LOCUS				
DEFINITION				
AC009602	37782 bp	DNA	INV	11-OCT-2000
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
	Leishmania major.			
	Leishmania major			
	Eukaryote; Euglenozoa; Kinetoplastida; Trypanosomatidae; ,			
	Leishmania.			
REFERENCE				
AUTHORS				
	1 (bases 1 to 37782)			
	Wyller,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.			
	Cawthra,J., Mairsolini,F., Sunkin,S. and Stuart,K.D.			
TITLE				
JOURNAL				
	Direct Submission			
	Submitted (28-AUG-1999) Seattle Biomedical Research Institution,			
	Nickerson Street, Seattle, WA 98109-1651, USA			

REFERENCE 2 (bases 1 to 37782)
AUTHORS Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,
McDonagh,P., Stuart,K. and Ivens,A.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2000) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
3 (bases 1 to 37782)
AUTHORS Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,
McDonagh,P., Stuart,K., Ivens,A. and Worthy,E.A.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-2000) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
COMMENT NOTE: This sequence is still preliminary and may contain some
errors, such as substitutions and frameshifts. Please regard all
CDS sequences with some degree of caution. A finished version
should be available within a few weeks.
FEATURES
source Location/Qualifiers
1..37782
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="35"
/clone="L2259"
complement(41..4125)
/gene="L2259.1"
/note="orf35.0650; predicted using Testcode and
Codonusage. ORF begins at 4125, Testcode and Codonusage
predict that 5th ATG at 3996 is used as start codon"
complement(41..3993)
/gene="L2259.1"
/note="Blastp similarity to MAP kinases in a number of
different organisms; Blastx similarity to PRP8 protein
homologs and splicing factors in a number of different
organisms;Eukaryotic protein kinase domain, TPR domain and
HCT-domain (ubiquitin-transferase) identified by Pfam;
identified as belonging to the family of Serine/threonine
protein kinases by COGS; HECT domain profile, Protein
kinase domain profile, TPR repeat region, circular profile
and TPR repeat profile identified by PROSITE;
Serine/threonine protein kinases active-site signature
pattern identified by PROSITE; Tyrosine protein kinases
specific active-site signature identified by Gribskov;
dileucine motif and coiled-coil region identified by
PSORT."
/codon_start=1
/product="L2259.1"
/protein_id="MAG22839.1"
/db_xref="GI:10765025"
/translation="MPLSMLOMGPSLVLSKRGVSVSAGPTRGCAEASPMAPDAOS
ASPGSVSGGVPMWVIOREQRIPLQFSATGSPNDGRLRREGNEARKAGYHEAI
RYTQAIEVDPSSEFLYTNRSFAVENIKFEKSAADAARAVELNANFPGHRLGIAO
MSLNDGHAMESLRKAMALAPSENKEAIRAMAKCSKMARAVPPLMSGGSSVP
YETGVSRSWPTAPRPTDAPALGISTVDCPELAEIRHAAATLRDAAYHRC
ADDAIVECSKRSRDKTSLSINSTSSLSLSEKADKOSQLRAVVDQSNAYHRC
DDDATLSKLMDTAQAGMAIDQKLTAKEEQNFSEFGASNALSINVEGTSIDLTL
AAAPAVVASPPASAPVPPNEAPADAGAGTSLSRASCISPGSSPTADAVAPHOPLT
TORAGHNHTOSTTAGALDQLRRRVEINBAVRGACSCFATDAAGAFVLSLAE
AEAMEEVRSLAEAVELDTQLETHARVOASTTKELSDVRLDALDOMSSIVIEVKRD
EAFVTRREGDGLLEEELKEORLLHERORLOLAEIMFKRIDPEPKIKITQYK
KRIHQRIAIHEKOKAVQTRIMELVEQDIPELAKSMANGSRILRVKSGSLMONLSF
SPFOYLSLSTVNSKVYHALRGEHVAKEISIDDAARRRQREVNIVTCNHPNI
IRIKGVFGPPAYLILPYRHGSLRALSKQSPMSVAVODMLROLASGVAYLHGHG
IYHGDLPKSNVLIADGPRVILDFGIADKDGALGVADMTLTNTVTNSGYSIVGTI
OYVAPQOLCTGSHSTAKSTGSDMVALGFTMLEVLALENAPFHDAISGKPSILLPE
OKRIDPAKAVGDEKLAEITIASVADPNARATADLLAHPEFSTLSSMNSQNS
AAKSDERIDAVRSYITAVRSRQVLYSVSRNMVESVRDIFORLDDDLSPYAV
VQGEAGIDEGALITTEMLTEYEQLLIVKALVCASSEAGAASPPASVLSSTAA
ASASAPNEGHSIVSESTVYLPAPDAGIADLPFLCKVLKLSILENRQIPLQJSS
SVLKPSGAEPSEFVLEEDYDPSIATSLRLRLSLAADLADAGIDFSEYTHFEFKAAD
GRYTDSPLTPEPSVDYISLRKRPDLVERRRTRALPMARKGFCSEPSLDHNLKLSPSD
LLILGOLHVASQYVDALEFGQFGRRNSNTPKYLKEVLLDMSQNNLRFLDLCTATA
AVPSGAMKKT"

gene complement(6308..7390)
/gene="L2259.2"
/note="orf35.0651; predicted using Glimmer, Testcode and
Codonusage"
CDS complement(6308..7390)
/gene="L2259.2"
/note="Blastp similarity to mitochondrial carrier proteins
and hypothetical mitochondrial proteins from a number of
different organisms; mitochondrial carrier protein motif
identified by Pfam; dileucine motif identified by PSORT;
Mitochondrial energy transfer proteins signature
identified by both PROSITE and Gribskov."
/codon_start=1
/product="L2259.2"
/protein_id="MAG22840.1"
/db_xref="GI:10765026"
/translation="MEVCSALAGIFARLCHPLDITKITYPTFGAGDHFAAPTAAS
SRTPAPTASFLOSARSIMAREGVAGYRRVGAALAGSAGVALISTYWCSTANKAF
GDGAAGAAGMAGTGESSPVGYMNTIQLPRLRAAATPSSVFVCGLAETVSCVW
VPIDVAKERLQSQPSPSLKGRYTSLDALRIIYNEGARGLYKYSTLSFGPFSVY
FVVEYETNMLAGLYTAASAPTCGDRNGSKERPSLATFVALGACAGQWVASIL
TNPLELVKTRIQORAVLHRDQGSTPALFSYHVRGLREGIALAKKEGIWALMKCV
GSRIATYAPNALTMGEFFELKTTLM"
complement(15577..17967)
/gene="L2259.3"
/note="orf35.0652; predicted using Glimmer, Testcode and
Codonusage"
complement(15577..17967)
/gene="L2259.3"
/note="nuclear localization predicted by PSORT;
Glutamine-rich, glycine-rich and proline-rich regions
identified by PROSITE; 7 transmembrane receptor (rhodopsin
family), Eukaryotic aspartyl protease, Bowman-Birk serine
protease inhibitor family."
Glutamate/leucine/phenylalanine/Valine dehydrogenase,
kinasin motor domain, Neurotransmitter-gated ion channel,
Orcidine 5'-phosphate decarboxylases, Oxidoreductases,
nltrogenase component 1 and other families,
phosphoglycerate kinases, Site-specific recombinases and
Regulator of G protein signalling domains identified by
Pfam; coiled-coil region identified by PSORT."
/codon_start=1
/product="L2259.3"
/protein_id="MAG22841.1"
/db_xref="GI:10765027"
/translation="MPSVRYTWYRRELRILATLASAMDGPVLRKPDVIEVAPVPT
PKRRAESNFKGSVLTDFNDSSTTTITNGNCGSGCGGSGNSGSSNNPGR
SSTPYVSGSGGCGDNRKRGCGGNGGDDSSNSVNSQSTYDRAPBEGFPNNPGR
NOETEEGLEYELRQSAIQRKVAETMERDEKRGELROLKLEKFOEGDAEAAAEK
EDDEIERLAGITIVDDAPKAEVRNHFESAQPTATSAHPATTOPSPPQASTLLS
FGAPASSIPATTTGTPQASVILPTPANSCTGQYARQPSWMSKSDSLMSTAPSMASALK
OSLOHSILPADASPSPVIRKPPQASQSQSPSCPASTSHQATSMCAAAANASATGQ
PASGASGCVAPAPASPPAPPOSVSGPAGHSGSTNTATMCKAKTQVSSAAQPA
PHTPPTQTHNSQSPSRAPATATIKARSSGSGTVPFGANSADLELLKKGATMI
SSAPAPSPSTASATIAKPTOPTIAELESMLMQARRQGVTKGMSLTVPSPPQFO
QPPRPAPQALLFANPSKTPRPANSPNTPSAPLSPQOLPASKMAMPNGMLQPKQPHQ
QOQPLRPPLRPMPRAHQSPQSPQOPQOPQOIOPOQOHPMMAMPRPQAPAP
QNSNSPMHNTPLPQIAPLAKMPQAPQPMQONCLOPMQONNAGQGVFVDAQLOARYMT
GQPMRPRRDGTVFMQAHMPQPOQAPRQQAPRPPNPGTMAQFLFQOQOQQRK"
complement(25479..27959)
/gene="L2259.4"
/note="orf35.0653; predicted using Glimmer, Testcode and
Codonusage. ORF begins at 27959, Testcode and Codonusage
predict that 3rd ATG at 27629 is used as start codon"
complement(25479..27629)
/gene="L2259.4"
/note="nuclear localization predicted by PSORT."
/codon_start=1
/product="L2259.4"
/protein_id="MAG22842.1"
/db_xref="GI:10765028"
/translation="MAHANGSGSLKGMTESDMRRLHSRYTPRNVSLPEFKSIVSGCTV
HHYIATLAKDILKIMQELRDLYEHGELREASVLPRLNGKQKQELSRRAATAVRLIDRSGDG

THIS PAGE BLANK (USPTO)

CC sequence represents a shortened version of the promoter sequence shown in
CC T42851. These sequences are efficient promoters for greatly increasing
CC the expression of foreign genes in transformant rice and other plants.
XX
50 Sequence 173 BP; 21 A; 48 C; 47 G; 57 T; 0 other;

Query Match	71.1%	Score 172;	DB 17;	Length 173;
Best Local Similarity	100.0%	Pred. No. 2.6e-42;		
Matches 172; Conservative	0;	Mismatches	0;	Gaps 0;

QY	71	taagcccaagctggtcttagagcctaaacgacacagaacctcttcgcgcgtctgtctctccag	130
Db	2	taagcccaagctggtcttagagcctaaacgacacagaacctcttcgcgcgtctgtctctccag	61
QY	131	ctcgaatcgtctctctgctctgctctgcctagaacctcctctgcctcgagctgcgtg	190
Db	62	ctcgaatcgtctctctgctctgctctgcctagaacctcctcctgcctcgagctgcgtg	121
QY	191	cttcgcctctccctcccaagctcgaatctgaatctgtgtgtgctggggggggcgag	242
Db	122	cttcgcctctccctcccaagctcgaatctgaatctgtgtgtgtgctggggggggcgag	173

RESULT	2
V05554	
ID	V05554 standard; DNA; 1.73 BP.

AC V05554;

DT 18-JUN-1998 (first entry)

Ubiquitin gene intron.

KW Intron; *Oryza sativa*; ubiquitous gene; ds.

Oryza sativa.

PN WO9747755-A1.

PD 18-DEC-1997.

PF 12-JUN-1997; 97WO-JP02030.

PR 12-JUN-1996; 96JP-0172922.

PA (NISB) JAPAN TOBACCO INC.

PI Kuraya Y, Morloka S, Ohta S, Ueki J;

DR WPT; 1998-086573/08.

PT Method for expressing foreign gene - comprises inserting gene

PT e.g. genetic engineering

PS Claim 4; Page 10; 20pp; Japanese.

CC This sequence represents an intron from the *Oryza sativa* ubiquitin gene.
CC and can be used in the method of the invention. The method is for
CC expressing a foreign gene by inserting the gene downstream from a
CC promoter, comprising at least 2 intron sequences. The method is used for
CC expression of foreign genes, useful in, e.g. genetic engineering. The
CC method achieves higher levels of expression than conventional methods.
XX
XQ Sequence 173 BP: 21 A: 48 C: 47 G: 57 T: 0 other:

Query Match	71.1%	Score 172;	DB 19;	Length 173;
Best Local Similarity	100.0%	Pred. NO.	2.6e-42;	
Matches 172; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY 71 taagccagtgtgcttaggctaagcgcaactagagcttcttgctcgcttgcttcttcgcg 130

Db 2 taagcccaagtgctctaaagcgaacctaagctcttcctgcctctctctccg 61
Oy 131 ctcaagatctgcttgcttgctctgcctgcctaaacctaactctgtgctgcgagatgcctg 190
Db 62 ctccagatctgcttgcttgctctgcctgcctaaacctaactctgtgctgcgagatgcctg 122
Oy 131 ctctgcctctctctctcaagctgcatactgatactgctgctgctggggggcgacg 242
Db 122 ctctgcctctctctctcaagctgcatactgatactgctgctgctggggggcgacg 173

RESULT	3
T42851	
ID	T42851 standard; DNA; 183 BP.

AC T42851;

DT 16-JUN-1997 (first entry)

DE Promoter #1 for phospholipase D gene.

KW Phospholipase D; r1ce; promoter; ss.

05 *Oryza sativa*.

PN W09630510-A1.

PD 03-OCT-1996.

PF 28-MAR-1996; 96WO-JP00812.

PR 29-MAR-1995; 95JP-0096126.

PA (NISB) JAPAN TOBACCO INC.

PI Morloka S, Ueki J;

DR WPT; 1996-455357/45.

PT Promoter DNA sequence derived from rice - used to increase

XX

XX
XX

CC are specifically promoters for the phospholipase D gene (PLD). T42852

CC efficient promoters for greatly increasing the expression of foreign

XX

Query Match	71.1%	Score 172:	DB: 17	Length 183;
Best Local Similarity	100.0%	Pred. No.	2.6e-42;	
Matches 172; Conservative	0;	Mismatches	0;	Gaps 0

Oy 71 taagccagatgtgttagagccaagcgcactatagatcttctgtctgtcttctctccg 130
 Db 7 taagccagatgtgttagagccaagcgcactatagatcttctgtctgtcttctccg 66
 Oy 131 ctgagatctgtctgtctgtctgtctgttagaacctactctgtgcgcgagtgctcg 190
 Db 67 ctcaagatctgtctgtctgtctgtctgttagaacctactctgtgcgcgagtgctcg 126
 Oy 191 ctctgtctctctctctcaagtcgcatctgattgtgtgtgtgtgtgtgtgtgtgtgag 242
 Db 127 ctctgtctctctctctcaagtcgcatctgattgtgtgtgtgtgtgtgtgtgtgtgag 178

RESULT 4
086785

ID	Q86785 standard; DNA: 2799 BP.
XX	
AC	Q86785;
DT	04-MAR-1996 (first entry)
XX	
DE	DNA encoding Phospholipase D.
XX	
KW	phospholipase D; measurement; reagent: phospholipid level; ss.
XX	
OS	Zea mays.
XX	
FM	key
FT	exon
FT	intron
FT	exon
FT	
FT	
FT	
PN	MO509234-A1.
XX	
PD	06-APR-1995.
XX	
PF	30-SEP-1994; 94WO-JP01627.
XX	
PR	30-SEP-1993; 93JP-026784.
XX	
PA	(NISB) JAPAN TOHACCO INC.
XX	
PI	Morioka S, Ueki J;
DR	WPI: 1995-147433/19.
DR	P-PSDB: R72799.
XX	
PT	Cloned DNA coding plant derived phospholipase D - controls
XX	
PS	expression of plant derived PLD gene
XX	
CC	Claim 14; Page 33-35; 41pp; Japanese.
XX	
CC	The DNA contains sequences necessary for the expression of a plant
CC	derived phospholipase D (PLD). The PLD is useful for measuring
CC	phospholipid levels and for producing derivs. by e.g. base exchange
CC	reactions.
XX	
SO	Sequence 2799 BP; 692 A; 709 C; 609 G; 789 T; 0 other:
<hr/>	
Query Match	71.1%; Score 172; DB 16; Length 2799;
Best Local Similarity	100.0%; Pred. No. 5,7e-42;
Matches 172; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	71 taagccacgagtgtcttaaggatacgcacatagaacttcttgctcgcttccttcocg 130
Db	1667 taagccacgagtgtcttaaggatacgcacatagaacttcttgctcgcttccttcocg 1726
OY	131 ctcaaatctgcttgcttgcttgcttcgctcgaaccctactctgtgctcgaaagtgcgctg 190
Db	1727 ctcaaatctgcttgcttgcttgcttcgctcgaaccctactctgtgctcgaaagtgcgctg 1786
OY	191 cttgctctctcctcctaagttcgattgtgtgtggggggcgacg 242
Db	1787 cttgctctcctcctccaagttcgattgtgtgtggggggcgacg 1838
<hr/>	
RESULT	5
ID	T42854
XX	T42854 standard; cDNA to mRNA: 2799 BP.
AC	T42854;
XX	
DT	16-JUN-1997 (first entry)

XX		Phospholipase D gene sequence fragment.
DE		
XX		Phospholipase D; rice; promoter; ss.
KW		
XX		
OS	Oryza sativa.	
XX		
FH	Key	Location/Qualifiers
FT	exon	1876..1983
FT		/tag= a
FT		/number= 1
FT		/note="encodes residues 1 to 36 of W06134"
FT	intron	1984..2523
FT		/tag= b
FT		/number= 1
FT	exon	2524..2799
FT		/tag= c
FT		/number= 2
FT		/note="encodes residues 37 to 128 of W06134"
PN		
XX	W0630510-A1.	
PD		
XX	03-OCT-1996.	
PX		
PF	28-MAR-1996;	96WO-JP00812.
XX		
PR	29-MAR-1995;	95JP-0096126.
XX		
PA	(NISB) JAPAN TOBACCO INC.	
XX		
PI	Morioka S, Ueki J;	
XX		
DR	WPI: 1996-455357/45.	
XX	P-PsDB; W06134.	
PT	Promoter DNA sequence derived from rice - used to increase expression of foreign genes in transformed hosts	
PS		
XX		
CC	Disclosure; Page 20-22; 29pp; Japanese.	
XX		
CC	This sequence represents a fragment of the coding sequence of the rice phospholipase D gene (PLD). The promoter for the PLD gene was isolated using the primers shown in T42857 and T42588. The promoters (see T42851 CC and T42852) are efficient promoters for greatly increasing the expression of foreign genes in transformant rice and other plants.	
CC		
SQ	Sequence 2799 BP; 692 A; 709 C; 609 G; 789 T; 0 other;	
Query Match	71.1%; Score 172; DB 17; Length 2799;	
Best Local Similarity	100.0%; Pred. No. 5.7e-42;	
Matches 172; Conservative	0; Mismatches 0; Indels 0; Gaps 0	
QY	71 taagccacgagtgtcttaaggctaagcgacataagagctcttgctgcgttctcttcgcg 130 	
Db	1667 taagccacgagtgtcttaaggctaagcgacataagagctcttgctgcgttctcttcgcg 1726	
QY	131 ctcaagatcgtcttgcttgcttgcttcgcgtagaacctactcgtgctgcgagtgcgtcg 190 	
Db	1727 ctcaagatcgtcttgcttgcttgcttcgcgtagaacctactcgtgctgcgagtgcgtcg 1786	
QY	191 ctctgcttccttcctccaagtcgaatcgtatgtgtgtgtggggggcgacag 242 	
Db	1787 ctctgcttccttcctccaagtcgaatcgtatgtgtgtgtggggggcgacag 1838	
RESULT	6	
ID	A10594/c	
AC	A10594 standard; DNA; 10732 BP.	
DT	29-JUN-2000 (first entry)	

```
XX DE Gene encoding a subunit of cellulose synthase.
XX KM Cellulose synthase; cellulose production; increase yield; ds.
XX OS Vigna angularis.
XX PN JP2000060568-A.
XX PD 29-FEB-2000.
XX PE 26-AUG-1998; 98JP-0239998.
XX PR 26-AUG-1998; 98JP-0239998.
XX PA (MIZU/) MIZUNO K.
XX PA (OJI) OJI PAPER CO.
XX DR WPI; 2000-342371/30.
XX DR P-PSDB; Y85179.
XX PT A gene encoding a cellulose synthetic equipment - for the improvement
XX PT in the amount of cellulose synthesised in a plant body
XX PS Claim 2; Page 14-21; 32pp; Japanese.
XX CC This sequence represents a gene encoding a subunit of the cellulose
XX CC synthase complex of Vigna angularis. The invention relates to subunits of
XX CC cellulose synthetic equipment, that can be used to increase the amount of
XX CC cellulose synthesised by a plant. The proteins and genes encoding them
XX CC can also be used to improve the properties of the cellulose being
XX CC produced by a plant.
XX SQ Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other:

Query Match 14.78; Score 35.6; DB 21; Length 10732;
Best Local Similarity 15.98; Pred. No. 0.2;
Matches 36; Conservative 99; Mismatches 92; Indels 0; Gaps 0;

OY 15 cttaggcgaagcgacatagatctctgtcgtcgtccttcctccgcgatccatcgaag 74
DB 9668 BTTGRTSTTYSYVSMCYDATTSRCTBYSTBRBTSSRCAKCTBDSTAKNSTSTYSSTRC 9609
OY 75 cccagtgctcttaggctaagcgcaactcctcgtcgtcgtcgtccttcctccgcga 134
DB 9608 TBYSRSRSGYVCRSRSRCYCTDSDSTCYSTTIVYSCCTTSKGYSDASRSTST 9549
OY 135 gatcgtcgtcgtcgtcgtcgtcgtcgaaccctactcgtcgtcgtcgtcgtcgtc 194
DB 9548 YSRCTTSTYSYSTDYSDCSSTTBNSTYSSDSDCTBYSSDRCSRSDSTCNCYSCSDS 9489
OY 195 gttctcctcctcgaagtcgatctgatgtgtgtgtg99999999999999999999 241
DB 9488 RYSTTYDACYYDAKTHCYTSDNCCNSTSRCTYNSTYSCSRBRSR 9442

RESULT 7
X13856
ID X13856 standard; DNA; 796 BP.
XX AC X13856;
XX DT 19-MAR-1999 (first entry)
XX DE Enterococcus faecalis genome contig SEQ ID NO:919.
XX KM Enterococcus faecalis; contig; detection; Enterococcal infection;
XX KM vaccine; attenuation; computer readable medium; ds.
XX OS Enterococcus faecalis.
XX PN WO9850555-A2.
```

```
XX PD 12-NOV-1998.
XX PF 04-MAY-1998; 98WO-US08985.
XX PR 14-NOV-1997; 97US-0066009.
XX PR 06-MAY-1997; 97US-0044031.
XX PR 16-MAY-1997; 97US-0046655.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Barash SC, Dillon PJ, Kunsch CA;
XX DR WPI; 1999-045171/04.
XX PT New isolated Enterococcus faecalis polynucleotides and polypeptides
XX PT - used to develop products for the detection of Enterococcus
XX PT use in vaccines for prevention or attenuation of Enterococcus
XX PT infection.
XX PS Claim 1; Page 2047-2048; 2084pp; English.
XX CC A computer readable medium has been developed which has recorded on it
XX CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
XX CC X12938 to X13919 represent these nucleotide sequences which are primary
XX CC nucleotide sequences, also known as contigs. The computer-based system
XX CC can identify fragments of the Enterococcus faecalis genome with
XX CC commercial importance. The products can be used to detect the presence
XX CC of Enterococcus faecalis in samples. They can also be used for
XX CC diagnosing Enterococcal infection in an animal and monitoring
XX CC progression of disease, and for identifying agents which can be used to
XX CC modulate the growth or pathogenicity of Enterococcus faecalis, or
XX CC another related organism, in vivo or in vitro. In particular the
XX CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
XX CC can be used in vaccines to prevent or attenuate an Enterococcal
XX CC infection.
XX SQ Sequence 796 BP; 113 A; 170 C; 163 G; 346 T; 4 other:

Query Match 14.18; Score 34.2; DB 20; Length 796;
Best Local Similarity 51.78; Pred. No. 0.25;
Matches 78; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

OY 63 tcagatcctaaagccagtgtagctgaagcgacgaagcgtcttcgtcgttgc 122
DB 368 tcacgcgcgaagcagtgtagctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtt 427
OY 123 ctctccgctcaagatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcga 182
DB 428 gatctgcacagtgtagctgttcaatgctcgtcgtcgtcgtcgtcgtcgtcgtcgtc 487
OY 183 tgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 213
DB 488 tgtagctcttctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 518

RESULT 8
T45666
ID T45666 standard; DNA; 2903 BP.
XX AC T45666;
XX DT 07-FEB-1997 (first entry)
XX DE Partial yeast gene, Lg-FLO1, involved in flocculation.
XX KM Flocculation; aggregation; floating; yeast; beer; Saccharomyces;
XX KM fermentation; ds.
XX OS Saccharomyces cerevisiae.
XX PN Key Location/Qualifiers
```

FT CDS 1..2550
FT /*tag= a
FT /note= "partial sequence only, no start codon"
PM JP08205900-A.
XX 13-AUG-1996.
PD 01-FEB-1995; 95JP-0015453.
XX 01-FEB-1995; 95JP-0015453.
XX 01-FEB-1995; 95JP-0015453.
XX (KIRI) KIRIN BREWERY KK.
XX WPI: 1996-419830/42.
DR P-PSDB; W06725.
XX
PT DNA molecule involved in yeast flocculation - useful for determining
PT If a sample yeast has a flocculation activity
XX
PS Claim 1; Page 6-10; 16pp; Japanese.
CC T45666 is a partial gene sequence from *Saccharomyces cerevisiae*. The
CC gene was named Ig-FLO1. The FLO1 protein is believed to be involved
CC in yeast flocculation (aggregation). The DNA and protein are used in
CC a method for determining whether or not a sample yeast has
CC flocculating activity, i.e. to distinguish if a yeast is a
CC flocculating or floating yeast.
SQ Sequence 2903 BP; 851 A; 776 C; 445 G; 831 T; 0 other;

Query Match 14.0%; Score 33.8; DB 17; Length 2903;
Best Local Similarity 54.4%; Pred. No. 0.47;
Matches 68; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

OY 37 tcttgcgcgtcttcttcctcgcagatcctcagccagtgcttaagcgaagc 96
DB 660 tctcttcagcactctctctcagcactcctcagtaattcttcctcagcacttc 719
OY 97 cactagagctcttgcgtcttcttcctcgcagatcctcagatgtcttgccttc 156
DB 720 tctctcagcactcctcagtaattcttcctcagcactcctcctatcactcttc 779
OY 157 gctag 161
DB 780 tccag 784

RESULT 9
X20580
ID X20580 standard; DNA: 19142 BP.
XX
AC X20580;
XX
DT 05-MAY-1999 (first entry)
XX
DE Polynucleotide sequence from the genome of *Treponema pallidum*.
XX
KW *Treponema pallidum* infection; syphilis; *Borrelia* infection; animal;
KW enzyme production; ds.
XX
OS *Treponema pallidum*.
XX
PN W09859034-A2.
XX
PD 30-DEC-1998.
XX
PE 23-JUN-1998; 98WO-US13041.
XX
PR 24-JUN-1997; 97US-0050667.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX
PI Fraser CM;
XX
DR WPI: 1999-081273/07.
PT New isolated *Treponema pallidum* nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of *T. pallidum* infections, particularly syphilis
XX
XX
PS Claim 1; Page 608-619; 1150pp; English.
CC X20500-21243 represent polynucleotide sequences from the genome of
CC *Treponema pallidum*. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for *T. pallidum*
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to *Borrelia* infections in animals, and for the
CC production of biosynthetic products such as enzymes.
XX
SQ Sequence 19142 BP; 4629 A; 5539 C; 4716 G; 4238 T; 20 other;

Query Match 13.7%; Score 33.2; DB 20; Length 19142;
Best Local Similarity 53.0%; Pred. No. 1.2;
Matches 71; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

OY 78 agtgccttaagcgaagcgaactagagctcttgcgtcgtctctctcgcgcagat 137
DB 3529 agagtcctcaggtccctccaattccgaactcgtcgtcgtctcttcttgccaagtgagc 3588
OY 138 ctgccttgccttgcgtcgtcgaaccctactctgtcgtcgaagtgctcgtcgtc 197
DB 3589 ctgcgtcgcactcgtcgtcttgaagaacttctgtatctgcacatgacatgcgcttttg 3648
OY 198 ttccttcctcaagt 211
DB 3649 ctgcctcactgaact 3662

RESULT 10
X40185
ID X40185 standard; DNA: 1095 BP.
XX
AC X40185;
XX
DT 02-JUL-1999 (first entry)
XX
DE Lung cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN W09904265-A2.
XX
PD 28-JAN-1999.
XX
PE 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Cure A, O'hare M, Odata Y, Old LJ;
PI Pfeundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;

DR	WPI: 1999-132448/11.
XX	
PT	New isolated cancer associated nucleic acids and polypeptides -
PT	isolated using sera from cancer patients, used to develop products
PT	for the diagnosis, monitoring or treatment of cancers
XX	
PS	Claim 67; Page 742; 787pp; English.
XX	
CC	The invention relates to a method for diagnosing a disorder characterised
CC	by expression of a human cancer associated antigen precursor coded for by
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC	biological sample isolated from a subject with an agent that specifically
CC	binds to the NAM, an expression product or a fragment of an expression
CC	product complexed with an HLA molecule; and (b) determining the
CC	interaction between the agent and the NAM or the expression product as a
CC	determination of the disorder. The products and methods can be used in
CC	the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens.
CC	The invention provides nucleic acid sequences and encoded polypeptides
CC	which are cancer associated antigen precursors expressed in human breast
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC	lung cancer.
XX	
SO	Sequence 1095 BP; 254 A; 269 C; 113 G; 432 T; 27 other;
XX	
QY	Query Match 13.6%; Score 33; DB 20; Length 1095;
	Best Local Similarity 54.5%; Pred. No. 0.62;
	Matches 66; Conservative 0; Mismatches 55; Indels 0; Gaps 0.
Db	38 tcttctgcgtctgtctcttcctcgcctcagatcctcctcagccagtgtgcttaagcgc 97
	7 tcccttcctcctctctctctctctctctctctctctctctctctctctcctcctcctcaccacaaatcgc 66
QY	98 actaagactctctgtctgcgtctcctcctccgcctcagatcgtctgttgcctcg 157
Db	67 actgctcctcctgaccttcttctcctatctcaccagaactcgtctgtctgtctcct 126
QY	158 c 158
Db	127 c 127
RESULT 11	
N70435/c	
ID N70435	standard; cDNA: 1136 BP.
XX N70435:	
AC	
XX	
DT	05-APR-1991 (first entry)
XX	
DE	Sequence encoding Insulin-like growth factor 1B (IGF-1B).
XX	
KW	Growth promoter; lactation enhancer; cell proliferation; ss.
XX	
OS	Homo sapiens.
XX	
PN	EP229750-A.
XX	
PD	22-JUL-1987.
XX	
PF	06-JAN-1987; 87EP-0870001.
XX	
PR	20-NOV-1986; 86US-0929671.
XX	
PR	07-JAN-1986; 86US-0816662.
XX	
PA	(UNIW) UNIV OF WASHINGTON.
XX	
PI	Krtvl GG, Rotweln PS;
XX	
DR	WPI: 1987-200203/29.
XX	
PT	New pre-pro-insulin-like growth factor-1 protein - obtd. by

PT	recombinant DNA procedures for use as growth promoters for
PT	enhancing lactation, for stimulating cell proliferation etc.
PS	Example; Fig 5; 59pp; English.
CC	A 42 base oligonucleotide corresponding to the DNA sequence encoding
CC	amino acids 10 to 23 of mature human IGF-I was synthesized (N70437).
CC	The radiolabeled 42 mer was then employed to screen for IGF-I
CC	containing DNA sequences in a human liver cDNA library. Insulin-
CC	like growth factors-1A and -1B cDNAs were isolated from a human cDNA
CC	library by using lambdaclat 11 (N70435, N70436). The human IGF-1
CC	genomic gene was isolated and mapped. It encodes at least two
CC	proinsulin-like growth factor-1 proteins. An essentially pure
CC	proinsulin-like growth factor-1 protein comprising the sequence
CC	of amino acids shown in Figure six is claimed (P70277).
XX	
XX	Sequence 1136 BP; 412 A; 230 C; 268 G; 226 T; 0 other:
XX	
XX	Query Match 13.6%; Score 32.8; DB 8; Length 1136;
XX	Best Local Similarity 52.1%; Pred. No. 0.71;
XX	Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY	103 agctcttgctgcgcgtctgtcttcttcctcgcctcagactctgcttgcctgcctaga 162
DB	727 ATCTCCCTCCCTGCTGCTCTTCTCTTCTCTCTGATCTGACAGCTGCTGCTCCCTCC
QY	163 accctactctgtgctgcgagctgcgctgcctgccttccttcctcctcaagctcagctgatt 222
DB	667 TTCTGTTCCCTCCCTGATGTCGTCTTCTGCGCCACCTTCTCTCTGAGACTTCTGTTTC 608
QY	223 gtctgtgtggtggggggtgcag 242
DB	607 TTGTTGTTAGATGGGGGCTG 588
XX	
XX	RESULT 12
XX	NS0149
XX	NS0149 standard; cDNA; 3336 BP.
XX	NS0149;
XX	AC
XX	27-SEP-1991 (first entry)
XX	DT
XX	Sequence of the cDNA clone 23B6p8.3 encoding a polypeptide
DE	exhibiting mammalian immunoglobulin binding factor activity (IBF).
XX	XX
KW	Immunoglobulin E-mediated disease; therapy; B-cell differentiation;
XX	immunoglobulin E-binding factor; ss.
XX	XX
XX	Key Location/Qualifiers
FT	CDS 94..1767
FT	/*tag= a
XX	
XX	EP155192-A.
XX	PN
XX	18-SEP-1985.
XX	PD
XX	15-MAR-1985; 85EP-0301834.
XX	PF
XX	16-MAR-1984; 84US-0590430.
XX	PR
XX	(SCHE-) SCHERING BIOTECH CO.
XX	(UYJO) JOHNS HOPKINS UNIV.
XX	PA (SCHE-) SCHERING-BIOTECH CO.
XX	PA (DNAX-) DNAX RES INST MOLEC.
XX	PI Martens CL, Ishizaka K, Moore KW, Huff TF;
XX	WP1, 1985-231863/38.
XX	P-PSDB: P50120.
XX	New complementary DNA clones coding for poly(peptide)s - with

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2001, 21:25:16 ; Search time 84.31 Seconds

(without alignments)
462.588 Million cell updates/sec

Title: US-09-600-602-2

Perfect score: 242

Sequence: 1 gtaagccagtgcttaagg.....gtgtgtgtggtgggggcacag 242

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues
Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/PCrus.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	172	71.1	173	1	US-08-750-007-7
2	172	71.1	183	1	US-08-750-007-1
3	172	71.1	2799	1	US-08-446-794A-5
4	172	71.1	2799	1	US-08-750-007-4
5	44.8	18.5	7218	3	US-08-232-463-14
6	30.8	12.7	6755	3	US-08-931-999-4
7	29	12.0	3531	2	US-08-714-402-1
8	28.4	11.7	16442	3	US-08-781-881-208
9	28.2	11.7	216	1	US-07-973-321B-1
10	28.2	11.7	216	1	US-08-090-527A-1
11	28.2	11.7	222	5	5274075-1
12	28.2	11.7	378	5	5274075-3
13	27.4	11.3	1950	1	US-08-592-126-93
14	27.2	11.2	3337	1	US-08-072-610-1
15	27.2	11.2	3337	2	US-08-719-822B-1
16	27	11.2	687	2	US-08-975-316-87
17	27	11.2	1178	2	US-08-107-676-26
18	27	11.2	1211	2	US-08-107-676-2
19	27	11.2	2277	1	US-08-676-967-5
20	27	11.2	2277	1	US-08-676-974-5
21	27	11.2	2277	2	US-09-096-487-5
22	26.8	11.1	688	2	US-08-975-316-88
23	26.8	11.1	1084	1	US-08-123-343A-4
24	26.8	11.1	2563	1	US-08-423-441-1
25	26.8	11.1	3900	1	US-08-123-343A-6
26	26.4	10.9	21126	1	US-08-008-216-19
27	26.4	10.9	21126	1	US-08-459-569-19
28	26.4	10.9	21126	1	US-08-458-831-19

C 29	26.4	10.9	43795	3	US-08-742-185-101	Sequence 101, App
C 30	26.2	10.8	1269	1	US-08-212-190A-1	Sequence 1, Appli
C 31	26.2	10.8	1269	2	US-08-900-321-1	Sequence 1, Appli
C 32	26.2	10.8	1269	1	PCT-US95-03610-1	Sequence 1, Appli
C 33	26.2	10.8	6453	1	US-08-306-691B-14	Sequence 14, Appli
C 34	26.2	10.8	6453	3	US-09-209-668-10	Sequence 10, Appli
C 35	26.2	10.8	6453	3	US-09-356-952-8	Sequence 8, Appli
C 36	26.2	10.8	9045	3	US-09-121-321-1	Sequence 1, Appli
C 37	26	10.7	1298	3	US-08-948-705-3	Sequence 8, Appli
C 38	26	10.7	1583	3	US-08-976-255-8	Sequence 8, Appli
C 39	26	10.7	1834	1	US-08-592-126-90	Sequence 90, Appli
C 40	25.8	10.7	4322	2	US-08-537-342-1	Sequence 1, Appli
C 41	25.8	10.7	9636	1	US-08-323-170B-1	Sequence 1, Appli
C 42	25.8	10.7	53526	3	US-08-658-136-2	Sequence 2, Appli
C 43	25.8	10.7	53577	3	US-08-658-136-1	Sequence 1, Appli
C 44	25.6	10.6	1175	1	US-07-791-377-1	Sequence 1, Appli
C 45	25.6	10.6	1490	3	US-08-605-150A-5	Sequence 5, Appli

ALIGNMENTS

```
RESULT 1
US-08-750-007-7
; Sequence 7, Application US/08750007
; Patent No. 5801016
; GENERAL INFORMATION:
; APPLICANT: MORIOKA, SHINJI
; APPLICANT: Ueki, JUN
; TITLE OF INVENTION: DNA FRAGMENT, RECOMBINANT VECTOR
; TITLE OF INVENTION: CONTAINING THE SAME AND METHOD FOR EXPRESSING FOREIGN
; NUMBER OF INVENTION: GENES USING THE SAME
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750, 007
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0760-221P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 173 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-750-007-7
```

```
Query Match 71.1%; Score 172; DB 1; Length 173;
Best Local Similarity 100.0%; Pred. No. 8.3e-45;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 71 taagccagtgcttaaggctaagcgcacctaagcttctgctgcttctcttcg 130
DB 2 TAAAGCCAGTGTGCTTAGGCTAAGCGCACCTAGAGCTTCTTGTGCTGCTTCTTCCG 61
```

Query Match	Best Local Similarity	Score	DB 1	Length	DB 2	Score	DB 3	Length
71	71.1%	172	1	183	130	71.1%	172	183
7	71.1%	172	1	183	130	71.1%	172	183
131	71.1%	172	1	183	130	71.1%	172	183
67	71.1%	172	1	183	130	71.1%	172	183
191	71.1%	172	1	183	130	71.1%	172	183
127	71.1%	172	1	183	130	71.1%	172	183

```

: Sequence 5 Application US/08446794A
: Patent No. 5747327
:
: GENERAL INFORMATION:
: APPLICANT: Ueki, Jun
: APPLICANT: MORIOKA, SHINJI
: TITLE OF INVENTION: PHOSPHOLIPASE D GENE ORIGINATED FROM
: TITLE OF INVENTION: PLANT
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
: STREET: P.O. BOX 747
: CITY: FALLS CHURCH
: STATE: VA
: COUNTRY: USA
: ZIP: 22040-0747
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/446,794A
: FILING DATE:
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: MURPHY JR, GERRALD M
: REGISTRATION NUMBER: 28,977
: REFERENCE/DOCKET NUMBER: 0760-0203P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-205-8000
: TELEFAX: 703-205-8050
:
: INFORMATION FOR SEQ ID NO: 5:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2799 base pairs
: type: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1876..1983
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2524..2799
:
: US-08-446-794A-5
:
: Query Match 71.1%; Score 172; DB 1; Length 2799;
: Best Local Similarity 100.0%; Pred. No. 2.3e-44;
: Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
: QY 71 taagccagctgcttagagcacaagcagcagcagcctctctgctgctctctctccg 130
: |||||||
: Db 1667 TAAGCCAGCTGTGCTTAGGCTAAGCAGCAGTACGACTTCTGCTGCTCTCTCCG 1726
:
: QY 131 ctcaagatcgtctgctgctgctgctgcgctagaaacctactctgctgcgagtgctg 190
: |||||||
: Db 1727 CTCGAGATCTGCTCTGCTTCGCTTCGCTAGAACCTACCTGTCGTCGAGTCTCG 1786
:
: QY 191 ctctgctctctctctctcaagtcgactgctgctgctgctgctgctgctgctg 242
: |||||||
: Db 1787 CTTCGTCTCTCTCTCAAGTTCGATCTGATTTGTTGTGTGGGGGGCGCAG 1838
:
: RESULT 4
: US-08-750-007-4
: Sequence 4, Application US/08750007
: Patent No. 5801016
: GENERAL INFORMATION:
: APPLICANT: MORIOKA, SHINJI
: APPLICANT: Ueki, Jun
: TITLE OF INVENTION: DNA FRAGMENT, RECOMBINANT VECTOR
: TITLE OF INVENTION: CLONING THE SAME AND METHOD FOR EXPRESSING FOREIGN

```



```

;
;   NUMBER OF SEQUENCES: 12
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/515,604
;   FILING DATE: 27-APR-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 468,766
;   FILING DATE: 23-JAN-1990
;   APPLICATION NUMBER: 468,766
;   FILING DATE: 23-JAN-1990
;   APPLICATION NUMBER: 369,625
;   FILING DATE: 21-JUN-1989
;   APPLICATION NUMBER: 272,243
;   FILING DATE: 16-NOV-1988
;   APPLICATION NUMBER: 229,178
;   FILING DATE: 05-AUG-1988
;   APPLICATION NUMBER: 226,421
;   FILING DATE: 29-JUL-1988
;   APPLICATION NUMBER: 140,036
;   FILING DATE: 31-DEC-1987
;
; SEQ ID NO:1
;
;   LENGTH: 222
5274075-1
```

```

Query Match          11.7%; Score 28.2; DB 5; Length 222;
Best Local Similarity 53.1%; Pred. No. 2.7;
Matches 60; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
```

```

QY 19 ggcgaagcgacactagagctctctgctgctctctccgcagatccgaagccca 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46 ggcgaagcgacactagagctctctgctgctctctccgcagatccgaagccca 105
QY 79 gtcgtcttaagcgaagcgaactgctctctgctgctctctctccgc 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106 gtgagctacagcgccgcctcaccgctcctcatggtgcagcgctccctcgaac 158
```

```

RESULT 12
5274075-3
;   APPLICANT: CHANG, TSE W.
;   TITLE OF INVENTION: NEWLY IDENTIFIED HUMAN EPSILON
;   IMMUNOGLOBULIN PEPTIDES AND RELATED PRODUCTS
;   NUMBER OF SEQUENCES: 12
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/515,604
;   FILING DATE: 27-APR-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 468,766
;   FILING DATE: 23-JAN-1990
;   APPLICATION NUMBER: 468,766
;   FILING DATE: 23-JAN-1990
;   APPLICATION NUMBER: 369,625
;   FILING DATE: 21-JUN-1989
;   APPLICATION NUMBER: 272,243
;   FILING DATE: 16-NOV-1988
;   APPLICATION NUMBER: 229,178
;   FILING DATE: 05-AUG-1988
;   APPLICATION NUMBER: 226,421
;   FILING DATE: 29-JUL-1988
;   APPLICATION NUMBER: 140,036
;   FILING DATE: 31-DEC-1987
;
; SEQ ID NO:3
;
;   LENGTH: 378
5274075-3
```

```

Query Match          11.7%; Score 28.2; DB 5; Length 378;
Best Local Similarity 53.1%; Pred. No. 3.3;
Matches 60; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
```

```

Db 202 ggcgaagcgccgctggaagctggaacgcgctctgcatcttcgcccgaactcttcgtcgaac 261
QY 79 gtcgtcttaagcgaagcgaactgctctctgctgctctctctccgc 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 262 gtgagctacagcgccgcctcaccgctcctcatggtgcagcggtccctcgaac 314
```

```

RESULT 13
US-08-592-126-93/c
; Sequence 93, Application US/08592126
; Patent No. 5821091
;
;   GENERAL INFORMATION:
;   APPLICANT: Gregory Dolganov
;   TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
;   OPERATING SYSTEM: Polypeptides
;   NUMBER OF SEQUENCES: 151
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Dehlinger & Associates
;   STREET: 350 Cambridge Avenue, Suite 250
;   CITY: Palo Alto
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94306
;
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patent In Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/592,126
;   FILING DATE:
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Sholtz, Charles K.
;   REGISTRATION NUMBER: 38,615
;   REFERENCE/DOCKET NUMBER: 4600-0111
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 324-0880
;   TELEFAX: (415) 324-0960
;   INFORMATION FOR SEQ ID NO: 93:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1950 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA to mRNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;   INDIVIDUAL ISOLATE: G310con.seq
;
; US-08-592-126-93
```

```

Query Match          11.3%; Score 27.4; DB 1; Length 1950;
Best Local Similarity 55.9%; Pred. No. 10;
Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```

QY 67 atccaaagcccaatgctgctgaagcgaagcgaactgctgctgctgctctc 126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 997 AGCGTGACCCACAGGACTTCAAGATAGGACGACTAGGCGCTTGAAGTCTTCTGC 938
QY 127 tccgctcagaatctgctgctgctgctgctcgc 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 937 CACAGGACATCCGGGATGATGATGATGATGCT 905
```

```

RESULT 14
US-08-072-610-1/c
; Sequence 1, Application US/08072610
; Patent No. 5531133
;
;   GENERAL INFORMATION:
;   APPLICANT: Barnwell, John
;   TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
```



```

; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072.610
; FILING DATE: 19930602
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/07686
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3337 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: PVMB3.3.1
; US-08-072-610-1

Query Match      11.2%; Score 27.2; DB 1; Length 3337;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 68; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 21 ctaagcgactagagctctctgctgctctctctcgcctcagatccttaagccagct 80
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2061 CTACGGTTCTCTTCTACTCTCTCTCTCGTGGTTCCTTAATTCGAATCGCTCTGGAG 2002

QY 81 ggccttagcgacgaagcactctctctgctgctctctctcgcctcagatctg 140
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2001 TTGCCTTAATTCCTCTCTGTGCGACTTCTCTGTGCTCTCTCTCTGTGCTCTT 1942

QY 141 ctgctgctgctgctc 156
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1941 CTTCGGTGGTCTCTCC 1926

RESULT 15
US-08-719-822B-1/C
; Sequence 1, Application US/08719822B
; Patent No. 5874527
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
```

```

; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719.822B
; FILING DATE: 09/30/96
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/17686US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3337 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: PVMB3.3.1
; US-08-719-822B-1

Query Match      11.2%; Score 27.2; DB 2; Length 3337;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 68; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 21 ctaagcgactagagctctctgctgctctctctcgcctcagatccttaagccagct 80
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2061 CTACGGTTCTCTTCTACTCTCTCTCTCGTGGTTCCTTAATTCGAATCGCTCTGGAG 2002

QY 81 ggccttagcgacgaagcactctctgctgctctctctcgcctcagatctg 140
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2001 TTGCCTTAATTCCTCTGTGCGACTTCTCTGTGCTCTCTCTCTGTGCTCTT 1942

QY 141 ctgctgctgctgctc 156
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1941 CTTCGGTGGTCTCTCC 1926
```

Search completed: February 12, 2001, 21:25:23
Job time: 3409 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

```
Run on:      February 12, 2001, 21:02:13 ; Search time 1923.53 seconds
              (without alignments)
              881.615 Million cell updates/sec
```

Title: US-09-600-602-2

Sequence: 1 gtaagccagtgcttag.....gtgtgtgtgtgggggcgcag 242

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues
Total number of hits satisfying chosen parameters: 15983484

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match 0%
	Maximum Match 100%
	Listing first 45 summaries

Database :

```

1:  EST: *
2:  1:  gb_est1: *
3:  2:  gb_est2: *
4:  3:  gb_est3: *
5:  4:  gb_est4: *
6:  5:  gb_est5: *
7:  6:  gb_est6: *
8:  7:  gb_est7: *
9:  8:  gb_est8: *
10: 9:  gb_est9: *
11: 10: gb_est10: *
12: 11: gb_est11: *
13: 12: gb_est12: *
14: 13: gb_est13: *
15: 14: gb_est14: *
16: 15: gb_est15: *
17: 16: gb_est16: *
18: 17: gb_est17: *
19: 18: gb_est18: *
20: 19: gb_est19: *
21: 20: gb_est20: *
22: 21: gb_est21: *
23: 22: gb_est22: *
24: 23: gb_est23: *
25: 24: gb_est24: *
26: 25: gb_est25: *
27: 26: gb_est26: *
28: 27: gb_est27: *
29: 28: gb_est28: *
30: 29: gb_est29: *
31: 30: gb_est30: *
32: 31: gb_est31: *
33: 32: gb_est32: *
34: 33: gb_est33: *
35: 34: gb_est34: *
36: 35: gb_est35: *
37: 36: gb_est36: *
38: 37: gb_est37: *
39: 38: gb_est38: *
40: 39: gb_est39: *
41: 40: gb_est40: *
42: 41: gb_est41: *
43: 42: gb_est42: *
44: 43: gb_est43: *
45: 44: gb_est44: *
46: 45: gb_est45: *
47: 46: gb_est46: *
48: 47: gb_est47: *
49: 48: gb_est48: *
50: 49: gb_est49: *
51: 50: gb_est50: *
52: 51: gb_est51: *
53: 52: gb_est52: *
54: 53: gb_est53: *
55: 54: gb_est54: *
56: 55: gb_est55: *
57: 56: gb_est56: *
58: 57: gb_est57: *
59: 58: gb_est58: *
60: 59: gb_est59: *
61: 60: gb_est60: *
62: 61: gb_est61: *
63: 62: gb_est62: *
64: 63: gb_est63: *
65: 64: gb_est64: *
66: 65: gb_est65: *
67: 66: gb_est66: *
68: 67: gb_est67: *
69: 68: gb_est68: *
70: 69: gb_est69: *
71: 70: gb_est70: *
72: 71: gb_est71: *
73: 72: gb_est72: *
74: 73: gb_est73: *
75: 74: gb_est74: *
76: 75: gb_est75: *
77: 76: gb_est76: *
78: 77: gb_est77: *
79: 78: gb_est78: *
80: 79: gb_est79: *
81: 80: gb_est80: *
82: 81: gb_est81: *
83: 82: gb_est82: *
84: 83: gb_est83: *
85: 84: gb_est84: *
86: 85: gb_est85: *
87: 86: gb_est86: *
88: 87: gb_est87: *
89: 88: gb_est88: *
90: 89: gb_est89: *
91: 90: gb_est90: *
92: 91: gb_est91: *
93: 92: gb_est92: *
94: 93: gb_est93: *
95: 94: gb_est94: *
96: 95: gb_est95: *
97: 96: gb_est96: *
98: 97: gb_est97: *
99: 98: gb_est98: *
100: 99: gb_est99: *
101: 100: gb_est100: *
102: 101: gb_est101: *
103: 102: gb_est102: *
104: 103: gb_est103: *
105: 104: gb_est104: *
106: 105: gb_est105: *
107: 106: gb_est106: *
108: 107: gb_est107: *
109: 108: gb_est108: *
110: 109: gb_est109: *
111: 110: gb_est110: *
112: 111: gb_est111: *
113: 112: gb_est112: *
114: 113: gb_est113: *
115: 114: gb_est114: *
116: 115: gb_est115: *
117: 116: gb_est116: *
118: 117: gb_est117: *
119: 118: gb_est118: *
120: 119: gb_est119: *
121: 120: gb_est120: *
122: 121: gb_est121: *
123: 122: gb_est122: *
124: 123: gb_est123: *
125: 124: gb_est124: *
126: 125: gb_est125: *
127: 126: gb_est126: *
128: 127: gb_est127: *
129: 128: gb_est128: *
130: 129: gb_est129: *
131: 130: gb_est130: *
132: 131: gb_est131: *
133: 132: gb_est132: *
134: 133: gb_est133: *
135: 134: gb_est134: *
136: 135: gb_est135: *
137: 136: gb_est136: *
138: 137: gb_est137: *
139: 138: gb_est138: *
140: 139: gb_est139: *
141: 140: gb_est140: *
142: 141: gb_est141: *
143: 142: gb_est142: *
144: 143: gb_est143: *
145: 144: gb_est144: *
146: 145: gb_est145: *
147: 146: gb_est146: *
148: 147: gb_est147: *
149: 148: gb_est148: *
150: 149: gb_est149: *
151: 150: gb_est150: *
152: 151: gb_est151: *
153: 152: gb_est152: *
154: 153: gb_est153: *
155: 154: gb_est154: *
156: 155: gb_est155: *
157: 156: gb_est156: *
158: 157: gb_est157: *
159: 158: gb_est158: *
160: 159: gb_est159: *
161: 160: gb_est160: *
162: 161: gb_est161: *
163: 162: gb_est162: *
164: 163: gb_est163: *
165: 164: gb_est164: *
166: 165: gb_est165: *
167: 166: gb_est166: *
168: 167: gb_est167: *
169: 168: gb_est168: *
170: 169: gb_est169: *
171: 170: gb_est170: *
172: 171: gb_est171: *
173: 172: gb_est172: *
174: 173: gb_est173: *
175: 174: gb_est174: *
176: 175: gb_est175: *
177: 176: gb_est176: *
178: 177: gb_est177: *
179: 178: gb_est178: *
180: 179: gb_est179: *
181: 180: gb_est180: *
182: 181: gb_est181: *
183: 182: gb_est182: *
184: 183: gb_est183: *
185: 184: gb_est184: *
186: 185: gb_est185: *
187: 186: gb_est186: *
188: 187: gb_est187: *
189: 188: gb_est188: *
190: 189: gb_est189: *
191: 190: gb_est190: *
192: 191: gb_est191: *
193: 192: gb_est192: *
194: 193: gb_est193: *
195: 194: gb_est194: *
196: 195: gb_est195: *
197: 196: gb_est196: *
198: 197: gb_est197: *
199: 198: gb_est198: *
200: 199: gb_est199: *
201: 200: gb_est200: *
202: 201: gb_est201: *
203: 202: gb_est202: *
204: 203: gb_est203: *
205: 204: gb_est204: *
206: 205: gb_est205: *
207: 206: gb_est206: *
208: 207: gb_est207: *
209: 208: gb_est208: *
210: 209: gb_est209: *
211: 210: gb_est210: *
212: 211: gb_est211: *
213: 212: gb_est212: *
214: 213: gb_est213: *
215: 214: gb_est214: *
216: 215: gb_est215: *
217: 216: gb_est216: *
218: 217: gb_est217: *
219: 218: gb_est218: *
220: 219: gb_est219: *
221: 220: gb_est220: *
222: 221: gb_est221: *
223: 222: gb_est222: *
224: 223: gb_est223: *
225: 224: gb_est224: *
226: 225: gb_est225: *
227: 226: gb_est226: *
228: 227: gb_est227: *
229: 228: gb_est228: *
230: 229: gb_est229: *
231: 230: gb_est230: *
232: 231: gb_est231: *
233: 232: gb_est232: *
234: 233: gb_est233: *
235: 234: gb_est234: *
236: 235: gb_est235: *
237: 236: gb_est236: *
238: 237: gb_est237: *
239: 238: gb_est238: *
240: 239: gb_est239: *
241: 240: gb_est240: *
242: 241: gb_est241: *
243: 242: gb_est242: *
244: 243: gb_est243: *
245: 244: gb_est244: *
246: 245: gb_est245: *
247: 246: gb_est246: *
248: 247: gb_est247: *
249: 248: gb_est248: *
250: 249: gb_est249: *
251: 250: gb_est250: *
252: 251: gb_est251: *
253: 252: gb_est252: *
254: 253: gb_est253: *
255: 254: gb_est254: *
256: 255: gb_est255: *
257: 256: gb_est256: *
258: 257: gb_est257: *
259: 258: gb_est258: *
260: 
```

44:	em_esthum2.*
45:	em_esthum3.*
46:	em_esthum4.*
47:	em_esthum5.*
48:	em_esthum6.*
49:	em_esthum7.*
50:	em_esthum8.*
51:	em_esthum9.*
52:	em_esthum10.*
53:	em_esthum11.*
54:	em_esthum12.*
55:	em_esthum13.*
56:	em_esthum14.*
57:	em_esthum15.*
58:	em_esthum16.*
59:	em_esthum17.*
60:	em_esthum18.*
61:	em_esthum19.*
62:	em_esthum20.*
63:	em_estln1.*
64:	em_estln2.*
65:	em_estln3.*
66:	em_estln4.*
67:	em_estov1.*
68:	em_estov2.*
69:	em_estp11.*
70:	em_estp12.*
71:	em_estp13.*
72:	em_estp14.*
73:	em_estp15.*
74:	em_estro1.*
75:	em_estro2.*
76:	em_estro3.*
77:	em_estro4.*
78:	em_estro5.*
79:	em_estro6.*
80:	em_estro7.*
81:	em_estro8.*
82:	em_estro9.*
83:	em_estro10.*
84:	em_estro11.*
85:	em_estro12.*
86:	em_estro13.*
87:	gb_estl1.*
88:	gb_estl2.*
89:	gb_estl3.*
90:	gb_estl4.*
91:	gb_estl5.*
92:	gb_estl6.*
93:	gb_estl7.*
94:	gb_estl8.*
95:	gb_estl9.*
96:	gb_estl10.*
97:	gb_estl11.*
98:	gb_estl12.*
99:	gb_estl13.*
100:	gb_estl14.*
101:	gb_estl15.*
102:	gb_estl16.*
103:	gb_estl17.*
104:	gb_estl18.*
105:	gb_estl19.*
106:	gb_estl20.*
107:	gb_estl21.*
108:	gb_estl22.*
109:	gb_estl23.*
110:	gb_estl24.*
111:	gb_estl25.*
112:	em_esthum21.*
113:	em_esthum22.*
114:	em_esthum23.*
115:	em_esthum24.*
116:	em_esthum25.*

117: em_estp16:*
118: em_estp17:*
119: em_estp18:*
120: em_estp19:*
121: em_estp20:*
122: em_estp21:*
123: em_estp22:*
124: em_estp23:*
125: em_estp24:*
126: em_estp25:*
127: em_estp26:*
128: em_estp27:*
129: em_estp28:*
130: em_estp29:*
131: em_estp30:*
132: em_estp31:*
133: em_estp32:*
134: em_estp33:*
135: em_estp34:*
136: em_estp35:*
137: em_estp36:*
138: em_estp37:*
139: em_estp38:*
140: em_estp39:*
141: em_estp40:*
142: em_estp41:*
143: em_estp42:*
144: em_estp43:*
145: em_estp44:*
146: em_estp45:*
147: em_estp46:*
148: em_estp47:*
149: em_estp48:*
150: em_estp49:*
151: em_estp50:*
152: em_estp51:*
153: em_estp52:*
154: em_estp53:*
155: em_estp54:*
156: em_estp55:*
157: em_estp56:*
158: em_estp57:*
159: em_estp58:*
160: em_estp59:*
161: em_estp60:*
162: em_estp61:*
163: em_estp62:*
164: em_estp63:*
165: em_estp64:*
166: em_estp65:*
167: em_estp66:*
168: em_estp67:*
169: em_estp68:*
170: em_estp69:*
171: em_estp70:*
172: em_estp71:*
173: em_estp72:*
174: em_estp73:*
175: em_estp74:*
176: em_estp75:*
177: em_estp76:*
178: em_estp77:*
179: em_estp78:*
180: em_estp79:*
181: em_estp80:*
182: em_estp81:*
183: em_estp82:*
184: em_estp83:*
185: em_estp84:*
186: em_estp85:*
187: em_estp86:*
188: em_estp87:*
189: em_estp88:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	42.6	17.6	1201	190	CNS0167A	AL106384 Drosophila
2	37.2	15.4	668	94	AM694929	AM694929 NF081B10S
3	37.2	15.4	671	94	AM694702	AM694702 NF079B05S
4	36.8	15.2	733	139	BF036458	BF036458 601460209
5	36.2	15.0	1078	190	CNS0004P	AL075925 Drosophila
6	36	14.9	997	190	CNS005TE	AL060767 Drosophila
7	35.8	14.8	852	191	CNS02F31	AL194518 Tetradon
8	35.8	14.8	974	137	BE880847	BE880847 601493284
9	35.4	14.6	897	94	AM728159	AM728159 CA_pa001
10	35.2	14.5	737	174	AZ031415	AZ031415 RPT-23-2
11	35.2	14.5	746	107	BE415895	BE415895 MUG002.B0
12	35	14.5	404	143	N61793	N61793 T9ESTy2891
13	35	14.5	743	109	BE519618	BE519618 HV_CEB001
14	35	14.5	1235	137	RE881061	RE881061 601492122
15	34.8	14.4	490	22	A1563663	A1563663 vY01a11.x
16	34.8	14.4	937	190	CNS006ST	AL065880 Drosophila
17	34.4	14.2	632	97	AM974965	AM974965 EST387070
18	34.4	14.2	846	106	BE287958	BE287958 601094475
19	34.4	14.2	872	176	A2184405	A2184405 SP_1002_B
20	34.2	14.1	352	13	AA920986	AA920986 vY16h09.r
21	34.2	14.1	715	87	AM174924	AM174924 f130c07.y
22	34.2	14.1	990	193	CNS04YVF	AL213476 Tetradon
23	33.8	14.0	441	9	AA562562	AA562562 v172a07.r
24	33.8	14.0	588	178	AZ337956	AZ337956 1M0068P13
25	33.8	14.0	595	172	AQ928561	AQ928561 RPT-23-2
26	33.8	14.0	702	106	BE268416	BE268416 601124712
27	33.8	14.0	915	135	BE798838	BE798838 601584668
28	33.8	14.0	1101	190	CNS017GS	AL108022 Drosophila
29	33.6	13.9	462	182	A2471215	A2471215 1M0285N10
30	33.6	13.9	488	110	BE638959	BE638959 946017C08
31	33.6	13.9	520	177	A2258776	A2258776 RPT-23-1
32	33.6	13.9	544	2	AA071714	AA071714 mm71c09.r
33	33.6	13.9	586	26	A1881474	A1881474 606069E10
34	33.6	13.9	586	109	BE553696	BE553696 ur41a10.x
35	33.6	13.9	659	182	A2463907	A2463907 1M0273K03
36	33.4	13.8	443	16	A1121302	A1121302 uc30D03.r
37	33.4	13.8	584	27	A1981330	A1981330 pat.pk005
38	33.4	13.8	1203	190	CNS015YR	AL106077 Drosophila
39	33.4	13.8	1265	139	BR032829	BR032829 601455658
40	33.2	13.7	1101	190	CNS016U2	AL107204 Drosophila
41	33.2	13.7	1101	190	CNS0173D	AL107539 Drosophila
42	33.2	13.7	1101	190	CNS0181N	AL108773 Drosophila
43	32.8	13.6	473	10	AA703492	AA703492 zj1h02.s
44	32.8	13.6	481	89	AM385721	AM385721 PM3-LT003
45	32.8	13.6	618	28	AJ398711	AJ398711 AJ398711

ALIGNMENTS

RESULT 1
CNS0167A
LOCUS
DEFINITION
CNS0167A 1201 bp DNA
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN15C02 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
AL106384
AL106384.1 GI:5621653
GSS.
GSS
26-JUL-1999

[illegible]

Plant Biology Division	
The Samuel Roberts Noble Foundation	
2510 Sam Noble Parkway, Ardmore, OK 73402, USA	
Tel: 580 221 7300	
Fax: 580 221 7380	
Email: radixon@noble.org	
Insert length: 668	Std Error: 0.00
Plate: 081	row: E column: 10
Seq primer: TCACACAGGAACACGATATGAC.	
Location/Qualifiers	
1..668	
/organism="Medicago truncatula"	
/db_xref="taxon:3880"	
/clone="NF081E10ST"	
/clone_lib="Developing stem"	
/tissue_type="stem"	
/dev_stage="Pooled developmental"	
/note="Vector: Lambda Zap; Contains a mixture of	
internodal stem segments"	
BASE COUNT	330 a 62 c 188 g 88 t
ORIGIN	
Query Match	15.4%; Score 37.2; DB 94; Length 668;
Best Local Similarity	59.4%; Pred. No. 2.9;
Matches	63; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
OY	117 ttgtctctctcccgctgacatcgtctctgtctgctgctgacgtagaccactctgtgc 176
Db	432 TTCCTTCGCATCGCTTCGCTCTGCTGCTCTTCTTTCATCTTCTTACTTATTTGT 373
OY	177 tgcgagtgtcgctgccttcgtctctctctcccaagtcgactcgtalt 222
Db	372 TTCACTTCTTCTTCATCTCTGCTTACTACTTTTCGTCGTAATT 327
RESULT	3
LOCUS	AM694702/c
DEFINITION	AM694702 671 bp mRNA EST 15-JUN-2000
VERSION	NF079805EM1F1044 Developing stem Medicago truncatula cDNA clone
KEYWORDS	NF079805ST 5', mRNA sequence.
ACCESSION	AM694702
SOURCE	AM694702.1 GI:7569464
ORGANISM	EST.
	barrel medic.
	Medicago truncatula
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
	Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE	1 (bases 1 to 671)
AUTHORS	He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
	,C.J., Flores,H.R., Imman,J.T., Weller,J.W., May,G.D. and Dixon
	,R.A.
TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation
JOURNAL	Medicago truncatula stem library
COMMENT	Unpublished (2000)
	Contact: Dixon RA
	Plant Biology Division
	The Samuel Roberts Noble Foundation
	2510 Sam Noble Parkway, Ardmore, OK 73402, USA
	Tel: 580 221 7302
	Fax: 580 221 7380
	Email: radixon@noble.org
	Insert length: 671
	Std Error: 0.00
	Plate: 079
	row: B column: 05
	Seq primer: TCACACAGGAACACGATATGAC.
	Location/Qualifiers
	1..671
FEATURES	/organism="Medicago truncatula"
SOURCE	/db_xref="taxon:3880"
	/clone="NF079805ST"
	/clone_lib="Developing stem"
	/tissue_type="stem"

		/dev_stage="Pooled developmental"
		/note="Vector: lambda Zap; Contains a mixture of
		Internodal stem segments"
BASE COUNT	314 a 80 c 180 g 97 t	
ORIGIN		
Query Match	15.4%:	Score 37.2;
Best Local Similarity	59.4%:	Pred. No. 2.9;
Matches	63;	Conservative 0; Mismatches 43; Indels 0; Gaps 0.
Dn	272 TTTCTTCCCATTCGGCGGTTTGCTGTGCTGTTTCATTTCTTAATTACATAAATTOT 213	
Qy	117 ttgtcttcttcgcagatcgtgcttgcttgctcgtagaacctactctgtgc 176	
Dn	212 TTTCAGCTTGTTTTTCATTTCTTGCCTTACTCCTTTTTCCTCGGAATT 167	
RESULT	4	
BFO36458/c	BF036458	733 bp mRNA EST 10-OCT-2000
LOCUS	601460209F1 NIH_MGC_66 Homo sapiens cdna clone IMAGE:3863690 5'	
DEFINITION	mRNA sequence.	
ACCESSION	BF036458	
VERSION	BF036458.1 GI:10744499	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 733)	
AUTHORS	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL	Unpublished (1999)	
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: DCRD/DTP CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LHCMS91 row: a column: 03 High quality sequence stop: 685. Location/Qualifiers	
FEATURES	source	location/qualifiers
	.. 733	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:3863690"	
	/clone_lib="NIH_MGC_66"	
	/tissue_type="adenocarcinoma"	
	/lab_host="DH10B (phage-resistant)"	
	/note="Organ: ovary; Vector: pCMV-SPORT6, site_1: NotI; site_2: SalI; Cloned unidirectionally. Primer: oligo dt. Average insert size 1.8 kb. Library constructed by Life Technologies."	
BASE COUNT	175 a 195 c 208 g 155 t	
ORIGIN		
Query Match	15.2%:	Score 36.8; DB 139; Length 733;
Best Local Similarity	61.5%:	Pred. No. 3.7;
Matches	59;	Conservative 0; Mismatches 37; Indels 0; Gaps 0.
Dn	729 AGAGGAGGCACGGGTTTTCCTCCTCGGCACACTTGAGGCTTGCTGCCCTTCCCCT 670	
Qy	101 agaagtcttcgtcgctgcttcttccgcgcccaaatcgtctgcttgctggcta 160	
Dn	161 gaacctactctgtgctgacgagtgctgctgttgtgt 196	

Db	669	GACCCCTCCTTCCTCCGCGTGTGCCTTTTAGGT	634
RESULT	5		
CNS00J4P			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
OY			
Dh			
OY			
Dh			
OY			
Dh			
OY			
Dh			
RESULT	6		
CNS00STE			
LOCUS			
DEFINITION			

ACCESSION	ALD060767
VERSION	ALD060767.1
KEYWORDS	GI:4943573
SOURCE	GSS.
ORGANISM	fruit fly. <i>Drosophila melanogaster</i> Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 997) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the <i>Drosophila melanogaster</i> genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library was prepared by Kazutoyo Oseguwa and Aaron Mammaster in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of <i>Drosophila</i> DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .

FEATURES	location/Qualifiers	
SOURCE	1..997	
	/organism="Drosophila	melanogaster"
	/db_xref="taxon:7227"	
	/clone_lib="RRC1_98"	
	/clone="BCR12K2"	
	/note="end : TET3"	
BASE COUNT	89 a 99 c 13 g	258 t 538 others
ORIGIN		

	ORIGIN
Query Match	14.9%; Score 36; DB 190; Length 997;
Best Local Similarity	20.3%; Pred No. 6.5;
Matches	39; Conservative 74; Mismatches 79; Indels 0; Gaps 0

[illegible]

RESULT	7
CNS02F31	
LOCUS	
DEFINITION	CNS02F31 852 bp DNA
	Tetradon nigroviridis genome survey sequence T7 end of clone
	11M13 of library G from Tetradon nigroviridis, genomic survey
	sequence.
ACCESSION	AL194518
VERSION	AL194518.1 GI:7832624
KEYWORDS	GSS: genome survey sequence.
SOURCE	Tetradon nigroviridis.
ORGANISM	Tetradon nigroviridis
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
1	Bouneau, L., Billault, A., Quettier, F., Sautin, W., Bernot, A. and Weissenbach, J.	Characterization and repeat analysis of the compact genome of the freshwater pufferfish <i>Tetraodon nigroviridis</i>	Unpublished	2	(bases 1 to 852)	Reest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quettier, F., Sautin, W. and Weissenbach, J.	Human gene number estimate provided by genome wide analysis using <i>Tetraodon nigroviridis</i> DNA sequence
3	Unpublished			3	(bases 1 to 852)	Genoscope.	Direct Submissions
4	Submitted (12-APR-2000)	to the EMBL/Genbank/DBJ databases		5	This sequence is a single read and was generated as part of a large scale clone-and sequencing project of the <i>tetraodon nigroviridis</i> genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .	Location/Qualifiers	1. 852

FEATURES	SOURCE	LOCATION/Qualifiers
	1. .852	
	/organism="Tetraodon nigrilividis"	
	/db_xref="taxon:99883"	
	/clone="13IN13"	
	/clone_11b="G"	
	/note="genoscope sequence ID : COAG313CG07LPI-end : T7"	
BASE COUNT	ORIGIN	
48 a	202 c	177 g 337 t 88 others

Query Match	14.8%	Score 35.8	DB 191	Length 852
Best Local Similarity	39.7%	Pred No. 7.1		
Matches 62	Conservative 25	Mismatches 69	Indels 0	Gaps 0

Oy	53	tccttcgcgcacgaagctcaagccacgaagtcgctcagaagcagcattagacgtctgc	112
	55		114
	57		116
	59		118
	61		120
	63		122
	65		124
	67		126
	69		128
	71		130
	73		132
	75		134
	77		136
	79		138
	81		140
	83		142
	85		144
	87		146
	89		148
	91		150
	93		152
	95		154
	97		156
	99		158
	101		160
	103		162
	105		164
	107		166
	109		168
	111		170
	113		172
	115		174
	117		176
	119		178
	121		180
	123		182
	125		184
	127		186
	129		188
	131		190
	133		192
	135		194
	137		196
	139		198
	141		200
	143		202
	145		204
	147		206
	149		208
	151		210
	153		212
	155		214
	157		216
	159		218
	161		220
	163		222
	165		224
	167		226
	169		228
	171		230
	173		232
	175		234
	177		236
	179		238
	181		240
	183		242
	185		244
	187		246
	189		248
	191		250
	193		252
	195		254
	197		256
	199		258
	201		260
	203		262
	205		264
	207		266
	209		268
	211		270
	213		272
	215		274
	217		276
	219		278
	221		280
	223		282
	225		284
	227		286
	229		288
	231		290
	233		292
	235		294
	237		296
	239		298
	241		300
	243		302
	245		304
	247		306
	249		308
	251		310
	253		312
	255		314
	257		316
	259		318
	261		320
	263		322
	265		324
	267		326
	269		328
	271		330
	273		332
	275		334
	277		336
	279		338
	281		340
	283		342
	285		344
	287		346
	289		348
	291		350
	293		352
	295		354
	297		356
	299		3

RESULT	8			
LOCUS	BE880847/c			
DEFINITION	BE880847	974 bp	mRNA	EST
ACCESSION	60149328	BE880847	BE880847	27-SEP-2000
VERSION	1	GI:10329623	EST	IMAGE:3895365.5'
KEYWORDS	BE880847.1	GI:10329623	EST	
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 974)			
AUTHORS	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ .			
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov			

DB	120	GCGTCGTGCTTGGTGATAGCCGCCTCCGAGACC	86
RESULT	13		
LOCUS	BE519618		
DEFINITION	BE519618	743 bp mRNA	EST 08-AUG-2000
		HV_CEBD0016H02f Hordeum vulgare seedling green leaf EST library	
		HVCNDM0005 (Erysiphe infected & control) Hordeum vulgare cDNA clone	
		HV_CEBD0016H02f, mRNA sequence.	
ACCESSION	BE519618		
VERSION	BE519618.1	GI:9743722	
KEYWORDS	EST.		
SOURCE	barley.		
ORGANISM	Hordeum vulgare		
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
	Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.		
REFERENCE	1 (bases 1 to 743)		
AUTHORS	Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Fritsch,D., Yu,		
	'Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo		
	'R., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D., and		
	Wood,T.		
TITLE	Development of a genetically and physically anchored EST resource		
JOURNAL	for barley genomics		
COMMENT	Unpublished (2000)		
	Contact: Wing RA		
	Clemson University Genomics Institute		
	Clemson University		
	100 Jordan Hall, Clemson, SC 29634, USA		
	Tel: 864 656 7288		
	Fax: 864 656 4293		
	Email: twing@clemson.edu		
FEATURES	Seq primer: AATTAACTCCTCAATAAAGG		
	High quality sequence start: 148		
	High quality sequence stop: 584.		
	Location/Qualifiers		
source	1..743		
	/organism="Hordeum vulgare"		
	/cultivar="CI16151 (MLab)"		
	/db_xref="taxon:4513"		
	/clone="HV_CEBD0016H02f"		
	/clone_lib="Hordeum vulgare seedling green leaf EST		
	library HVCNDM0005 (Erysiphe infected & control)"		
	/tissue_type="seedling green leaf"		
	/lab_host="SOLR"		
	/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	24 a 256 c 88 g 363 t 12 others		
ORIGIN			
Query Match	14.5%, Score 35; DB 109; Length 743;		
Best Local Similarity	55.3%; Pred. No. 12;		
Matches	68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;		
OY	105 ctcttgatgtcgtgccttcctctccgcagcatcgctgcttgctgcttcgcacgaac	164	
Db	392 CTGTGCTTTTCGCCGCTCTTCTCTGTTACTCCGTTGTCCTGTTCATGCTC	451	
OY	165 ccatactcgtgcgtagcgaagtcgcgcgtcgtccttcctccaagtgcgatgatgt	224	
Db	452 CCTGCTTTGCTCGTGCCTTCCTCTTTGATCGTCTTCTGTTCTGTTGCTTTT	511	
OY	225 gtg 227		
Db	512 TTG 514		
RESULT	14		
LOCUS	BE881061/c		
DEFINITION	BE881061 1235 bp mRNA	EST 27-SEP-2000	
	60149212JEL NIH_MGC_69 Homo sapiens cDNA IMAGE:3894292 5',		
	mRNA sequence.		
ACCESSION	BE881061		
VERSION	BE881061.1	GI:10329837	

[illegible]

THIS PAGE BLANK (USPTO)

22	34.2	19.9	154.959	52	AC0251125	AC0251125 Homo sapi
23	34.2	19.8	95280	53	AC0267770	AC0267770 Homo sapi
24	34.2	19.8	166680	42	AC020685	AC020685 Homo sapi
25	34	19.7	116534	58	AC069368	AC069368 Homo sapi
26	33.8	19.5	326695	50	AC022126	AC022126 Homo sapi
27	33.6	19.4	534	80	HUMUT1934	HUMUT1934 Homo sapi
28	33.6	19.4	82188	6	AB010076	AB010076 Arabidops
29	33.4	19.3	377	80	HUMUT846	HUMUT846 Homo sapi
30	33.4	19.3	39229	8	AC004755	AC004755 Homo sapi
31	33.4	19.3	163297	57	AC068847	AC068847 Homo sapi
32	33.2	19.2	1470	73	TRPBP	M17716 T.pallidum
33	33.2	19.2	14464	1	AE001269	AE001269 T.reponema
34	33.2	19.2	68418	65	AC079068	AC079068 Homo sapi
35	33.2	19.2	167994	39	AC020864	AC020864 Homo sapi
36	33.2	19.2	168853	38	AC011691	AC011691 Homo sapi
37	33.2	19.2	174487	44	AC012011	AC012011 Homo sapi
38	33.2	19.2	198211	59	AC073756	AC073756 Mus muscu
39	33.2	19.2	204504	41	AC016820	AC016820 Homo sapi
40	33.2	19.2	209920	59	AC075682	AC075682 Mus muscu
41	33.2	19.2	223193	51	AC024462	AC024462 Homo sapi
42	33	19.1	37782	31	AC098602	AC098602 Leishman
43	33	19.1	230760	59	AC073714	AC073714 Mus muscu
44	32.8	19.0	185	81	E01350	E01350 Synthetic T
45	32.8	19.0	185	81	I08008	I08008 Sequence 1

ALIGNMENTS

RESULT	1			
AR037064				
LOCUS	AR037064	173 bp	DNA	PAT
DEFINITION	Sequence 7 from patent US 5801016.			29-SEP-1999

VERSION	AR037064.1	GI:5954920		
KEYWORDS				
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 173)			
TITLE	Morioka, S. and Ueki, J.			
JOURNAL	DNA fragment, recombinant vector containing the same and method for			
FEATURES	expressing foreign genes using the same			
source	Patent: US 5801016-A 7 01-SEP-1998;			
	location/Qualifiers			
	1..173			
	/organism="unknown"			
BASE COUNT	21 a	48 c	47 g	57 t
ORIGIN				

Query Match	100.0%	Score 173;	DB 81;	Length 173;
Best Local Similarity	100.0%	Pred. No. 1.2e-42;		
Matches 173; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

RESULT	2				
AR037061					
LOCUS	AR037061	183 bp	DNA		
DEFINITION	Sequence 1 from patent US 5801016.			PAT	29-SEP-1999
ACCESSION	AR037061				

VERSION	AR037061.1	GI:5954917	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 183)		
TITLE	Morioka,S. and Ueki,J.		
JOURNAL	DNA fragment, recombinant vector containing the same and method for		
FEATURES	expressing foreign genes using the same		
SOURCE	Patent: US 5801016-A 1 01-SEP-1998;		
	Location/Qualifiers		
	1..183		
	/organism="unknown"		
BASE COUNT	23 a	51 c	51 g
ORIGIN			58 t

Query Match	100.0%	Score 173;	DB 81;	Length 183;
Best Local Similarly	100.0%	Pred. No. 1.2e-42;		
Matches 173; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	gtaagcccaagtgcttccttagctagctaaagcagcaatagagctctctgcagctctctctctcc	60
Db	6	gtaagcccaagtgcttccttagctagctaaagcagcaatagagctctctgcagctctctctctcc	65
QY	61	gctcaagatcctctgtcttgccttgcctcagaaaccttactctgtgttgagatgct	120
Db	66	gctcaagatcctctgtcttgccttgcctcagaaaccttactctgtgttgagatgct	125
QY	121	gcttcgctctctctctcctaagttcgaatcagattgtgtgtgagggggagcgag	173
Db	126	gcttcgctctctctctcctaagttcgaatcagattgtgtgtgagggggagcgag	178

[illegible]

	Query Match	Similarity	100.0%	Score 173:	DB 81:	Length 2799:
	Best Local	Similarity	100.0%	Pred. No. 1,3e-42:		
	Matches 173:	Conservative	0:	Mismatches 0:	Indels 0:	Gaps 0:
Qy	1	gtaaacccagatgctcttagagctgaagaagcacaagaactctctgcacgtgctctctcc	60			
Db	1666	GTAAccCAGtGCTTAgGCTAAcCGACtTAAGAcCTTCTTgcCTTCTTCTCC	1725			
Qy	61	gtcagacatctgcttgcttgcttgcttcgtcagaaacctactctgtgcgcgagatgcct	120			
Db	1726	GCTcAGATcTGCTTGCTTGCTTGCTTGCTGCAACCCtACTGtGCGtGAGAGTGCCT	1785			
Qy	121	gcttcgctctccctccctcaagttcgaatctgtatgtgtgtgtgtgtgtgtgtgtgtgtgt	173			
Db	1786	GCTTcGCTTCTCCCTTCACGTTCAGTTCATCTGATTGtGTGTGTGGGGGGGGCGAG	1838			

RESULT 4

[illegible]

```

RESULT      8
LOCUS       AB042235/c
DEFINITION Homo sapiens genomic DNA, chromosome 6q21, clone:419N8, complete
              sequence.
ACCESSION   AB042235
VERSION     AB042235.1
KEYWORDS    HTG.
SOURCE      Homo sapiens
            Homo sapiens DNA, clone:419N8.
REFERENCE   1 (bases 1 to 93540)
            Mammalia: Eutheria: Chordata: Craniata: Vertebrata: Euteleostomi;
            Human.
AUTHORS     Nakamura,Y. and Koyama,K.
JOURNAL     Human BAC clone 419N8
REFERENCE   2 (bases 1 to 93540)
            Published Only in Database (2000) In press
AUTHORS     Nakamura,Y. and Koyama,K.
JOURNAL     Direct Submission
            Submitted (23-APR-2000) to the DDBJ/EMBL/GenBank databases. Yusuke
            Nakamura, Institute of Medical Science, The University of Tokyo,
            Human Genome Center, Laboratory of Molecular Medicine; 4-6-1
            Shirokane, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:kumi@kokains.u-tokyo.ac.jp, Tel:01-3-5449-5373,
            Fax:01-3-5449-5406)
FEATURES
    source
        1..93540
            Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="6"
                /clone="419N8"
                /map="6q21"
BASE COUNT  30813 a 16408 c 16565 g 29754 t
ORIGIN

```

Query Match	20.5%;	Score 35.4;	DB 8;	Length 93540;
Best Local Similarity	55.2%;	Pred. No. 2.2;		

RESULT	9
AL133403/c	
LOCUS	AL133403 126334 bp DNA HTG 05-OCT-2000
DEFINITION	Homo sapiens chromosome 6 clone RP1-196P3, *** SEQUENCING IN PROGRESS ***, 13 unordered pieces.
ACCESSION	AL133403
VERSION	AL133403.10 GI:10715703
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Mumayyala, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eumariota; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 126334)
AUTHORS	Sims,S.
TITLE	Direct Submission
JOURNAL	Submitted (04-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT	Requests: clonerequests@sanger.ac.uk On Oct 6, 2000 this sequence version replaced gi:9926404.

```

Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: DJ1966P
Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 21% of reads
Sequencing vector: pLacI; L08752; 78% of reads
Chemistry: Dye-terminator Big Dye; 10% of reads
Consensus quality: 121560 bases at least Q40
Consensus quality: 123203 bases at least Q30
Consensus quality: 124166 bases at least Q20
Insert size: 125114; sum-of-contigs
Insert size: 266896; 21.2% error; agarose-fp
Quality coverage: 4.63x in Q20 bases; sum-of-contigs Quality
coverage: 2.28x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
8115: contig of 8115 bp in length
*
8116 8215: gap of 100 bp
*
8216 21939: contig of 13724 bp in length
*
21940 22039: gap of 100 bp
*
22040 30793: contig of 8753 bp in length
*
30793 30892: gap of 100 bp
*
30893 36737: contig of 5845 bp in length
*
36738 36837: gap of 100 bp
*
36838 45878: contig of 9041 bp in length
*
45879 45978: gap of 100 bp
*
45979 57041: contig of 11063 bp in length

```

*	57042	57141:	gap of	100 bp			
*	57142	65961:	contlg of	8840 bp	in	length	
*	65982	66081:	gap of	100 bp			
*	66082	72218:	contlg of	6137 bp	in	length	
*	72219	72318:	gap of	100 bp			
*	72319	74457:	contlg of	2179 bp	in	length	
*	74458	74597:	gap of	100 bp			
*	74598	81911:	contlg of	7314 bp	in	length	
*	81912	82011:	gap of	100 bp			
*	82012	84713:	contlg of	2702 bp	in	length	
*	84714	84813:	gap of	100 bp			
*	84814	87242:	contlg of	2429 bp	in	length	
*	87243	87342:	gap of	100 bp			
*	87343	126334:	contlg of	38992 bp	in	length	

FEATURES	source	Location/Qualifiers
		1. .126334
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="6"
		/clone="RP1-196P3"
		/clone_1lb="RPC1-1"
		1. .8115
misc_feature		/note="assembly_fragment:01345
		clone_end:SP6
		vector_side:left"
		8216. .21939
misc_feature		/note="assembly_fragment:00285
		fragment_chain:1"
		22040. .30792
misc_feature		/note="assembly_fragment:00821
		fragment_chain:1"
		30893. .36737
misc_feature		/note="assembly_fragment:00720
		fragment_chain:2"
		36838. .45878
misc_feature		/note="assembly_fragment:00454
		fragment_chain:2"
		45979. .57041
misc_feature		/note="assembly_fragment:00868
		fragment_chain:3"
		57142. .65981
misc_feature		/note="assembly_fragment:01575
		fragment_chain:3"
		66082. .72218
misc_feature		/note="assembly_fragment:00232"
		72319. .74497
misc_feature		/note="assembly_fragment:00778"
		74598. .81911
misc_feature		/note="assembly_fragment:00907.0"
		82012. .84713
misc_feature		/note="assembly_fragment:01052"
		84814. .87242
misc_feature		/note="assembly_fragment:01090"
		87343. .126334
misc_feature		/note="assembly_fragment:01341"
		21219 c 21563 g 40860 t 1203 others
BASE COUNT	41489 a	
ORIGIN		

	Query Match	20.5%	Score 35.4	DB 66	Length 126334	
	Best Local Similarity	55.2%	Pred. No. 2.2			
	Matches 69	Conservative 0	Mismatches 56	Indels 0	Gaps 0	
OY	13 tgccttagagcctaagcgactagaacgtcttcttgccgcttgtcttccttcgccgaatcgc	72				
Db	881 TTCTGTGTCCTCCCTCCTCCTCTTCTTCTATCCCTTTCTTCCTTCCTTCGCTT	822				
OY	73 ttgccttgcttgcttcgcctcagaacctactcgtgctcgcgagtcgcgcgtctgccttc	132				
Db	821 TTTCCTTCCTTCCTCCCTCCCTCCCTCTCTCTTCTTCCTTCCTTCCTTCCTTCT	762				
OY	133 ttcccl 137					

Db 761 TCTT 757

RESULT 10	AB041340/c		
LOCUS	AB041340/c		
DEFINITION	HOMO SAPIENS genomic DNA, chromosome 6q21, clone:553P24, complete sequence.	PRI	04-AUG-2000
ACCESSION	AB041340		
VERSION	AB041340.1	GI:9711362	
KEYWORDS	HTG.		
SOURCE	HOMO SAPIENS DNA, clone_11b:Research Genetics, human BAC library clone:553P24.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 128241)		
TITLE	Nakamura, Y. and Koyama, K.		
JOURNAL	Human BAC clone 553P24		
REFERENCE	Published only in Database (2000) In press		
AUTHORS	2 (bases 1 to 128241)		
TITLE	Nakamura, Y. and Koyama, K.		
JOURNAL	Direct Submission		
	Submitted (05-APR-2000) to the DDBJ/EMBL/GenBank databases. Yusuku Nakamura, Institute of Medical Science, The University of Tokyo, Human Genome Center, Laboratory of Molecular Medicine; 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan		
	(E-mail:kumikok@ims.u-tokyo.ac.jp, Tel:81-3-5449-5373, Fax:81-3-5449-5406)		

```
FEATURES
source
    location/Qualifiers
    1. 128241
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="6"
        /clone="53p24"
        /clone_lib="Research Genetics, human BAC Library"
        /map="6q21"
BASE COUNT
ORIGIN
41573 a 22014 c 21763 g 42851 t
```

Query Match:	20.5%	Score 35.4;	DB 8;	Length 128241;
Best Local Similarity:	55.2%	Pred No. 2.2;		
Matches 69;	Conservative 0;	Mismatches 56;	Indels 0;	Gaps 0

OY	13	ttcctagagctaagcgcactagaagctctctgtcgtctgtctctctccgcatacgtgc	72
Db	116238	ttctttgtttcttcttccctccctcttcttcttcttcttcttcttcttcttcttctt	116179
OY	73	ttctctgtctgtctgtcgtctagaacctctgtgtctgtcgaatgtcgtcgtctctcc	132
Db	116178	tttccttcttcttcttcttccctccctcccttcttcttcttcttcttcttcttcttctt	116119
OY	133	ttctct	137
Db	116118	ttcttt	116114

RESULT	11
LOCUS	AL355532/c
DEFINITION	AL355532 181302 bp DNA
	Human DNA sequence from clone RP11-487f5 on chromosome 6, complete sequence.
ACCESSION	AL355532
VERSION	AL355532.10
KEYWORDS	GI:10045412
SOURCE	HTG.
ORGANISM	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 181302)
AUTHORS	Cobley, V.
TITLE	Direct Submission

JOURNAL Submitted (07-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Sep 9, 2000 this sequence version replaced g1:9926594.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>
Rp11-487F5 is from the library RPCI-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://Dacpac.med.buffalo.edu/VECTOR.pbace3.6>

FEATURES

source

This sequence is the entire insert of clone Rp11-487F5 the true left end of clone Rp11-196P3 is at 166177 in this sequence. The true right end of clone Rp11-423D17 is at 30283 in this sequence.
Location/Qualifiers
1..181302
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="Rp11-487F5"
/clone_11b="RPCI-11.2"
165..209
repeat_region
/note="L1M1 repeat: matches 6139..6185 of consensus"
210..575
repeat_region
/note="PHE1B repeat: matches 1..364 of consensus"
576..1153
repeat_region
/note="L1M1 repeat: matches 5595..6139 of consensus"
1167..1267
repeat_region
/note="L1M8 repeat: matches 6075..6171 of consensus"
1268..1528
repeat_region
/note="AluY repeat: matches 35..295 of consensus"
1529..2819
repeat_region
/note="L1M8 repeat: matches 4748..6075 of consensus"
2940..4010
repeat_region
/note="L1M8 repeat: matches 3502..4589 of consensus"
4033..4216
repeat_region
/note="AluJo repeat: matches 3..170 of consensus"
4238..4259
repeat_region
/note="11 copies 2 mer to 100% conserved"
4264..4819
repeat_region
/note="L1M8 repeat: matches 2857..3448 of consensus"
4820..5020
repeat_region
/note="match: GSS: Em:A0752208"
4951..5032
repeat_region
/note="L1M4 repeat: matches 4517..4588 of consensus"
5032..5389
repeat_region
/note="L1P3 repeat: matches 5792..6150 of consensus"
5390..5803
repeat_region
/note="match: GSS: Em:A0286760"
5804..5815
repeat_region
/note="match: GSS: Em:A0745266"
5816..5825
repeat_region
/note="match: GSS: Em:A0884205"
5826..5872
repeat_region
/note="match: GSS: Em:A0432202"

misc_feature
complement(5498..5803)
/note="match: GSS: Em:A0269249"
5804..7766
repeat_region
/note="L1M1 repeat: matches 3572..5793 of consensus"
5807..5808
misc_feature
/note="match: GSS: Em:A0134658"
5809..8378
repeat_region
/note="L1M1 repeat: matches 3014..3565 of consensus"
8379..9420
repeat_region
/note="L1M2 repeat: matches 731..1693 of consensus"
9421..9713
repeat_region
/note="L1M2 repeat: matches 144..409 of consensus"
9714..10581
repeat_region
/note="L1M2 repeat: matches -669..-242 of consensus"
10582..11744
repeat_region
/note="L1M4 repeat: matches 2111..2829 of consensus"
11745..11934
repeat_region
/note="L1M4 repeat: matches 1888..2044 of consensus"
11935..12128
repeat_region
/note="AluY repeat: matches 128..311 of consensus"
12129..12167
repeat_region
/note="19 copies 2 mer to 78% conserved"
12168..12310
repeat_region
/note="AluY repeat: matches 1..132 of consensus"
12311..12456
repeat_region
/note="L1M4 repeat: matches 318..518 of consensus"
12457..12780
repeat_region
/note="L1M4 repeat: matches -177..-47 of consensus"
12781..12969
repeat_region
/note="L1M4 repeat: matches -269..-196 of consensus"
12970..13050
repeat_region
/note="MER54B repeat: matches 1..790 of consensus"
13051..17008
repeat_region
/note="L1M4 repeat: matches 209..464 of consensus"
17009..17242
repeat_region
/note="Char1ie2 repeat: matches 92..360 of consensus"
17243..18008
repeat_region
/note="Char1ie2 repeat: matches 3145..3672 of consensus"
18009..19124
repeat_region
/note="match: GSS: Em:A0371909"
19125..19345
repeat_region
/note="MER6B repeat: matches 74..486 of consensus"
19346..19412
repeat_region
/note="MER6A repeat: matches 1..79 of consensus"
19413..20224
repeat_region
/note="L1P16 repeat: matches 4667..6157 of consensus"
20225..21691
repeat_region
/note="match: GSS: Em:B36088"
21692..21957
repeat_region
/note="MER5A repeat: matches 40..178 of consensus"
21958..22909
repeat_region
/note="L2 repeat: matches 2616..2745 of consensus"
22910..23484
repeat_region
/note="L1M2 repeat: matches 1..397 of consensus"
23485..24316
repeat_region
/note="L2 repeat: matches 1242..1814 of consensus"
24317..24717
repeat_region
/note="match: GSS: Em:A0344549"
24718..24803
repeat_region
/note="match: GSS: Em:A0267518"
24804..24815
repeat_region
/note="match: STS: Em:G15813"
24816..25084
repeat_region
/note="48 copies 2 mer to 74% conserved"
25085..25382
repeat_region
/note="AluJo repeat: matches 3..295 of consensus"
25383..27981
repeat_region
/note="L1P16B repeat: matches 9..458 of consensus"
27982..28454
repeat_region
/note="L2 repeat: matches 192..362 of consensus"
28455..29766
repeat_region
/note="L2 repeat: matches 2648..2709 of consensus"


```

* 10540 12080: contig of 1541 bp in length
* 12081 12180: gap of 100 bp
* 12181 14702: contig of 2522 bp in length
* 14703 14802: gap of 100 bp
* 14803 16733: contig of 1931 bp in length
* 16734 16833: gap of 100 bp
* 16834 18755: contig of 1922 bp in length
* 18756 18855: gap of 100 bp
* 18856 20786: contig of 1931 bp in length
* 20787 20886: gap of 100 bp
* 20887 23125: contig of 2239 bp in length
* 23126 23225: gap of 100 bp
* 23226 25267: contig of 2042 bp in length
* 25268 25367: gap of 100 bp
* 25368 27294: contig of 1927 bp in length
* 27295 27394: gap of 100 bp
* 27395 29083: contig of 1689 bp in length
* 29084 29183: gap of 100 bp
* 29184 31810: contig of 2627 bp in length
* 31811 31910: gap of 100 bp
* 31911 34614: contig of 2704 bp in length
* 34615 34714: gap of 100 bp
* 34715 36824: contig of 2110 bp in length
* 36825 36924: gap of 100 bp
* 36925 39358: contig of 2434 bp in length
* 39359 39458: gap of 100 bp
* 39459 41675: contig of 2217 bp in length
* 41676 41775: gap of 100 bp
* 41776 44142: contig of 2367 bp in length
* 44143 44242: gap of 100 bp
* 44243 45373: contig of 1133 bp in length
* 45376 45475: gap of 100 bp
* 45476 48030: contig of 2555 bp in length
* 48031 48130: gap of 100 bp
* 48131 50762: contig of 2632 bp in length
* 50763 50862: gap of 100 bp
* 50863 54865: contig of 4003 bp in length
* 54866 54965: gap of 100 bp
* 54966 58786: contig of 3821 bp in length
* 58787 58886: gap of 100 bp
* 58887 62079: contig of 3193 bp in length
* 62080 62179: gap of 100 bp
* 62180 66768: contig of 4589 bp in length
* 66769 66868: gap of 100 bp
* 66869 70876: contig of 4008 bp in length
* 70877 70976: gap of 100 bp
* 70977 75996: contig of 5020 bp in length
* 75997 76096: gap of 100 bp
* 76097 80273: contig of 4177 bp in length
* 80274 80373: gap of 100 bp
* 80374 85679: contig of 5306 bp in length
* 85680 85779: gap of 100 bp
* 85780 91717: contig of 5538 bp in length
* 91718 91817: gap of 100 bp
* 91818 97046: contig of 5229 bp in length
* 97047 97146: gap of 100 bp
* 97147 101127: contig of 3981 bp in length
* 101128 101227: gap of 100 bp
* 101228 106788: contig of 5561 bp in length
* 106789 106888: gap of 100 bp
* 106889 111554: contig of 4666 bp in length
* 111555 111654: gap of 100 bp
* 111655 118598: contig of 6944 bp in length
* 118599 118698: gap of 100 bp
* 118699 124207: contig of 5509 bp in length
* 124208 124307: gap of 100 bp
* 124308 130324: contig of 6017 bp in length
* 130325 130424: gap of 100 bp
* 130425 138528: contig of 8104 bp in length
* 138529 138628: gap of 100 bp
* 138629 146991: contig of 8363 bp in length
* 146992 147091: gap of 100 bp
* 147092 153531: contig of 6440 bp in length

```

```

* 153532 153631: gap of 100 bp
* 153632 162631: contig of 9000 bp in length
* 162632 162731: gap of 100 bp
* 162732 174676: contig of 11945 bp in length
* 174677 174776: gap of 100 bp
* 174777 186910: contig of 12134 bp in length
* 186911 187010: gap of 100 bp
* 187011 201251: contig of 14241 bp in length.
FEATURES
  source
    location/Qualifiers
    1. 201251
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /chromosome="11"
    /map="11"
    /clone="RP23-136C7"
    /clone_id="RPI-23 Female Mouse BAC"
  misc_feature
    1. 1487
    /note="assembly-fragment"
  misc_feature
    1588. 2793
    /note="assembly-fragment"
  misc_feature
    2894. 4247
    /note="assembly-fragment"
  misc_feature
    4348. 5849
    /note="assembly-fragment"
  misc_feature
    5950. 7248
    /note="assembly-fragment"
  misc_feature
    7349. 8876
    /note="assembly-fragment"
  misc_feature
    8977. 10439
    /note="assembly-fragment"
  misc_feature
    10540. 12080
    /note="assembly-fragment"
  misc_feature
    12181. 14702
    /note="assembly-fragment"
  misc_feature
    14803. 16733
    /note="assembly-fragment"
  misc_feature
    16834. 18755
    /note="assembly-fragment"
  misc_feature
    18856. 20786
    /note="assembly-fragment"
  misc_feature
    20887. 23125
    /note="assembly-fragment"
Query Match
  Best Local Similarity 50.9%; Score 35.4; DB 52; Length 201251;
  Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
Oy 6 cccagtgctgaagcgaagcgaactagaagctctgctgctctctccgctca 65
Db 42131 CTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 42072
Oy 66 gatcgcctgctgctgctgctgctgctgctgctgctgctgctgctgctc 125
Db 42071 CTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 42012
Oy 126 gtctccctcccaagcgaatgctgctgctgctgctgctgctgctgctgct 170
Db 42011 CTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 41967
RESULT 13
AC026682
LOCUS
DEFINITION Mus musculus chromosome 11 clone RP23-201K2, WORKING DRAFT
ACCESSION AC026682
VERSION AC026682.9 GI:10645292
KEYWORDS HTG; HTGS; PHASE1; HTGS_DRAFT.
SOURCE
  ORGANISM Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 204806)

```

AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C., Dedecker, D., Thomas, S., Okunou, G., Carlack, C., Garner, T., Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J., Bunney, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z., Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S., Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R., Gorenfeld, J., H., Gunaratne, P., Haller, G., Hernandez, J., Hognes, M., Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S., Kover, C., Liu, J., Liu, W., Louisse, D., Lorado, R., Martin, R., Massey, E., McLeod, M., P., Mel, G., Moore, S., Morgan, M., Morris, S., Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B., Perez, L., Reller, D., Say, J., Shen, H., Vasquez, L., Watlington, S., Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A., Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G., Worley, K. and Gibbs, R.	Unpublished	Direct Submission	2 (bases 1 to 204806)	Worley, K. C.	Direct Submission	Submitted (23-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	On Oct 5, 2000 this sequence version replaced gi:9930744.
						Genome Center	
						Center: Baylor College of Medicine	
						Center code: BCM	
						Web site: http://www.hgsc.bcm.tmc.edu/	
						Contact: hgsc-help@bcm.tmc.edu	
						Project Information	
						Center project name: MADH	
						Center clone name: RP23-201K2	
						Summary Statistics	
						Sequencing vector: M13; L08821	
						Chemistry: Dye-terminator Big Dye; 74% of reads	
						Assembly program: Phrap; version 0.990329	
						Consensus quality: 204424 bases at least Q40	
						Consensus quality: 210404 bases at least Q30	
						Consensus quality: 212657 bases at least Q20	
						Estimated insert size: 203906; sum-of-contigs estimation	
						Quality coverage: 0x in Q20 bases; agarose-gel estimation	
						Quality coverage: 7x in Q20 bases; sum-of-contigs estimation	
						NOTE: Estimated insert size may differ from sequence length	
						(see http://www.hgsc.bcm.tmc.edu/docs/centbank_draft_data.html)	
						NOTE: This is a 'working draft' sequence. It currently	
						consists of 6 contigs. The true order of the pieces	
						is not known and their order in this sequence record is	
						arbitrary. Gaps between the contigs are represented as	
						runs of N, but the exact sizes of the gaps are unknown.	
						This record will be updated with the finished sequence	
						as soon as it is available and the accession number will	
						be preserved.	
						1 86612: contig of 86612 bp in length	
						86613 86712: gap of unknown length	
						86713 157534: contig of 70822 bp in length	
						157535 157634: gap of unknown length	
						157635 176497: contig of 18863 bp in length	
						176498 176597: gap of unknown length	
						176598 194222: contig of 17625 bp in length	
						194223 203398: gap of unknown length	
						203399 203498: contig of 9076 bp in length	
						203499 204806: gap of unknown length	
						204806: contig of 1308 bp in length.	
						Location/Qualifiers	
						1. 204806	
						/organism="Mus musculus"	
						/db_xref="taxon:10090"	
						/chromosome="11"	
						/clone="RP23-201K2"	
						55914 a 47687 c 46801 g 53899 t 515 others	
						BASE COUNT	
						ORIGIN	

	Query Match	Best Local Similarity	20.5%:	Score 35.4:	DB 53:	Length 204806;
			50.9%:	Pred. No. 2.3:		
	Matches 84:	Conservative	0:	Mismatches	81:	Indels 0: Gaps 0:
OY	6	cceagtgcttgaagctaagcgcacataagaacttcgttcgctgtcattcccgatca 65				
Db 109862		CTCCCTTCCTTCCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 109921				
OY	66	gatcgcgttcgttcgttcgttcgtagaacctactcgttcgtcgagtgcgtcttc 125				
Db 109922		CTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 109981				
OY	126	gtcttccttcctcaagttcatgcacatcatgttgttgtgtggggggagcg 170				
Db 109982		CCTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCG 110026				
RESULT 14						
AP002530/C						
LOCUS						
DEFINITION		Homo sapiens genomic DNA, chromosome 6q21, anti-oncogene region,				
		section 3/4.				
ACCESSION		AP002530				
VERSION		AP002530.2				
KEYWORDS		GI:9081804				
SOURCE		.				
ORGANISM		Homo sapiens DNA.				
		Homo sapiens				
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE						
AUTHORS		1 (sites)				
TITLE		Nakamura,Y. and Koyama,K.				
JOURNAL		DNA sequence analysis of a 1-Mb region on chromosome 6q21				
REFERENCE		Published Only in Database (2000) In press				
AUTHORS		2 (bases 1 to 300000)				
TITLE		Hirakawa,M. and Yamaguchi,H.				
JOURNAL		Direct Submission				
		Submitted (15-JUN-2000) to the DDBJ/EMBL/Genbank databases. Mka				
		Hirakawa, Japan Science and Technology Corporation (JST), Advanced				
		Databases Department; Yonbancho 5-3, Chiyoda-ku, Tokyo 102-0081,				
		Japan (E-mail:m.kake@tokyo.jst.go.jp,				
		url:http://www-alis.tokyo.jst.go.jp/, Tel.:81-3-5214-8491,				
		Fax:81-3-5214-8470)				
COMMENT						
		On Jul 12, 2000 this sequence version replaced gl:8570520.				
		This sequence is conducted by Japanese Foundation				
		for Cancer Research as a JST sequencing Team.				
		Principal Investigator: Yusuke Nakamura Ph.D				
		Phone:+81-3-5449-5372, Fax:+81-3-5449-5433,				
		yusuke@nc.lms.u-tokyo.ac.jp				
		The sequence is submitted by Human Genome Sequencing				
		in AIMS project of JST				
		Japan Science and Technology Corporation (JST)				
		5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan				
		For further information about this sequences,				
		please visit our sequence archive Web site				
		(http://www-alis.tokyo.jst.go.jp/HGS/top.html)				
		or send email to webmaster@www-alis.tokyo.jst.go.jp.				
FEATURES						
source						
		1..300000				
		/organism="Homo sapiens"				
		/db_xref="taxon:9606"				
		/chromosome="6"				
		/map="6q21"				
		82803..83089				
STS		/note="AFM190yfl0:The location is between each flanking				
		site of PCR primers."				
		/db_xref="GDB:188224"				
		240123..240372				
STS		/note="CHIC.ATA22H10.PI6079:The location is between each				
		flanking site of PCR primers."				
		/db_xref="GDB:683877"				
		240174..240320				

[illegible]

```

90 - 99 :      81266
-----
Percentage of bases with a quality value >= 40 : 96 %.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
    source
        1..201833
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="14"
            /clone_1id="RPCI-11"
            /clone="R-404P21"
BASE COUNT      51535 a 47502 c 46494 g 56301 t      1 others
ORIGIN
Query Match          20.2% ; Score 35; DB 84; Length 201833;
Best Local Similarly 53.2% ; Pred. No. 3;
Matches 74; Conservative 0; Mismatches 65; Indels 0; Gaps

OY       23   aagcgacactagaagcttcgtcgtcgtccatcttcctgcgctagatcgttgctgctt 82
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 146883 AAAGCAGATCATTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 146942

OY       83   gcttcgactagaaccctactcgtgctgcygaagtgcgctgcttcgcttcctcctaagt 142
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 146943 CCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 147002

OY       143   tcgatcgtatgtytgtygt 161
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 147003 TCTTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 147021

```

```
Search completed: February 12, 2001, 21:30:17
Job time: 4703 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2001, 21:27:54 : Search time 136.04 Seconds
(without alignments)
477.725 Million cell updates/sec

Title: US-09-600-602-3

Perfect score: 173
Sequence: 1 gtaagcccaagtgctgaag.....gtgtgtgtggtgggggagcgag 173

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_36:*

- 1: /SID6/gcgdata/geneseq/geneseqn/NA1980.DAT:*
- 2: /SID6/gcgdata/geneseq/geneseqn/NA1981.DAT:*
- 3: /SID6/gcgdata/geneseq/geneseqn/NA1982.DAT:*
- 4: /SID6/gcgdata/geneseq/geneseqn/NA1983.DAT:*
- 5: /SID6/gcgdata/geneseq/geneseqn/NA1984.DAT:*
- 6: /SID6/gcgdata/geneseq/geneseqn/NA1985.DAT:*
- 7: /SID6/gcgdata/geneseq/geneseqn/NA1986.DAT:*
- 8: /SID6/gcgdata/geneseq/geneseqn/NA1987.DAT:*
- 9: /SID6/gcgdata/geneseq/geneseqn/NA1988.DAT:*
- 10: /SID6/gcgdata/geneseq/geneseqn/NA1989.DAT:*
- 11: /SID6/gcgdata/geneseq/geneseqn/NA1990.DAT:*
- 12: /SID6/gcgdata/geneseq/geneseqn/NA1991.DAT:*
- 13: /SID6/gcgdata/geneseq/geneseqn/NA1992.DAT:*
- 14: /SID6/gcgdata/geneseq/geneseqn/NA1993.DAT:*
- 15: /SID6/gcgdata/geneseq/geneseqn/NA1994.DAT:*
- 16: /SID6/gcgdata/geneseq/geneseqn/NA1995.DAT:*
- 17: /SID6/gcgdata/geneseq/geneseqn/NA1996.DAT:*
- 18: /SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:*
- 19: /SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:*
- 20: /SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:*
- 21: /SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	173	100.0	173	17 T42852
2	173	100.0	173	17 V05554
3	173	100.0	183	17 T42851
4	173	100.0	2799	16 Q86785
5	173	100.0	2799	17 T42854
6	33.2	19.2	796	20 X13856
7	33.2	19.2	19142	20 X20580
8	32.8	19.0	1136	8 N70435
9	31.4	18.2	10732	21 A10594
10	30.8	17.8	6755	19 V21511
11	29.8	17.2	1954	20 X21686
12	29.8	17.2	3850	20 X21656

C 13	28.8	16.6	586	20	X20167
C 14	28.8	16.6	2559	20	X20162
C 15	28.8	16.6	2559	20	X20164
C 16	28.8	16.6	2559	20	X20166
C 17	28.8	16.6	4169	21	A09331
C 18	28.8	16.6	4594	20	X13167
C 19	28.8	16.6	2384	20	Z24428
C 20	28.6	16.4	11883	21	A10263
C 21	28.4	16.2	12036	11	O04668
C 22	28	16.2	109	20	Z32037
C 23	28	16.2	300	20	Z13115
C 24	28	16.2	700	20	V88907
C 25	28	16.2	876	20	Z16225
C 26	28	16.2	876	20	Z16226
C 27	28	16.2	1391	21	Z57866
C 28	27.6	16.0	1950	19	V60018
C 29	27.4	15.8	3531	20	X77451
C 30	27.2	15.7	1664976	19	V21209
C 31	27	15.6	1671	12	Q11799
C 32	27	15.6	2374	19	V05805
C 33	27	15.6	2374	19	V05807
C 34	27	15.6	114955	20	X53491
C 35	26.8	15.5	270	15	O66941
C 36	26.8	15.5	728	21	A02243
C 37	26.8	15.5	1084	15	O66946
C 38	26.8	15.5	2563	14	O50008
C 39	26.8	15.5	3900	15	O66950
C 40	26.6	15.4	368	20	V88624
C 41	26.6	15.4	1209	21	Z86988
C 42	26.6	15.4	6002	21	Z86988
C 43	26.6	15.4	162450	21	Z86967
C 44	26.4	15.3	944	20	X78475
C 45	26.4	15.3	1095	20	X40185

ALIGNMENTS

RESULT 1

ID	Sequence	Score	Query Match Length	ID	Description
T42852	T42852 standard; DNA; 173 BP.	173	100.0	173	17 T42852
AC	T42852:	173	100.0	173	17 V05554
XX	16-JUN-1997 (first entry)	173	100.0	173	17 T42851
DE	Promoter #2 for phospholipase D gene.	173	100.0	173	17 T42854
XX	Phospholipase D; rice; promoter; ss.	173	100.0	173	17 T42854
KW	Oryza sativa.	173	100.0	173	17 T42854
OS	Oryza sativa.	173	100.0	173	17 T42854
XX	Oryza sativa.	173	100.0	173	17 T42854
PN	WO9630510-A1.	173	100.0	173	17 T42854
XX	WO9630510-A1.	173	100.0	173	17 T42854
PD	03-OCT-1996.	173	100.0	173	17 T42854
XX	03-OCT-1996.	173	100.0	173	17 T42854
PF	28-MAR-1996; 96WO-JP00812.	173	100.0	173	17 T42854
XX	28-MAR-1996; 96WO-JP00812.	173	100.0	173	17 T42854
PR	29-MAR-1995; 95JP-0096126.	173	100.0	173	17 T42854
XX	29-MAR-1995; 95JP-0096126.	173	100.0	173	17 T42854
PA	(NISR) JAPAN TOBACCO INC.	173	100.0	173	17 T42854
XX	(NISR) JAPAN TOBACCO INC.	173	100.0	173	17 T42854
PI	Morioka S, Ueki J;	173	100.0	173	17 T42854
XX	Morioka S, Ueki J;	173	100.0	173	17 T42854
DR	WPI; 1996-455357/45.	173	100.0	173	17 T42854
XX	WPI; 1996-455357/45.	173	100.0	173	17 T42854
PT	Promoter DNA sequence derived from rice - used to increase	173	100.0	173	17 T42854
XX	Promoter DNA sequence derived from rice - used to increase	173	100.0	173	17 T42854
PS	expression of foreign genes in transformed hosts	173	100.0	173	17 T42854
XX	expression of foreign genes in transformed hosts	173	100.0	173	17 T42854
PS	Claim 3; Page 23; 29pp; Japanese.	173	100.0	173	17 T42854
XX	Claim 3; Page 23; 29pp; Japanese.	173	100.0	173	17 T42854
CC	T42851 and T42852 represent promoters isolated from rice. These sequences	173	100.0	173	17 T42854
XX	T42851 and T42852 represent promoters isolated from rice. These sequences	173	100.0	173	17 T42854
CC	are specifically promoters for the phospholipase D gene (PLD). This	173	100.0	173	17 T42854

CC sequence represents a shortened version of the promoter sequence shown in
CC T42851. These sequences are efficient promoters for greatly increasing
CC the expression of foreign genes in transformant rice and other plants.
XX

SO Sequence 173 BP; 21 A; 48 C; 47 G; 57 T; 0 other;

Query Match 100.0%; Score 173; DB 17; Length 173;
Best Local Similarity 100.0%; Pred. No. 8.8e-45;
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaagcccaagtgtgcttagagctaaagcgcactagagctcttgcgtctctctcc 60
|||||
DB 1 gtaagcccaagtgtgcttagagctaaagcgcactagagctcttgcgtctctctcc 60
61 gctcagatctgctgtgctgtgcttgcgtcagaacctactctgtgcgagtgctgct 120
61 gctcagatctgctgtgctgtgcttgcgtcagaacctactctgtgcgagtgctgct 120
QY 121 gctcgtctctctcctcccaagttcgatctgctgtgtggtggtggtggtggtggt 173
|||||
DB 121 gctcgtctctctcctcccaagttcgatctgctgtgtggtggtggtggtggtggt 173

RESULT 2

V05554 standard; DNA; 173 BP.

XX V05554;

XX 18-JUN-1998 (first entry)

XX Ubiquitin gene intron.

XX Intron; Oryza sativa; ubiquitin gene; ds.

XX Oryza sativa.

XX WO9747755-A1.

XX 18-DEC-1997.

XX 12-JUN-1997; 97WO-JP02030.

XX 12-JUN-1996; 96JP-0172922.

XX (NISR) JAPAN TOBACCO INC.

XX Kuraya Y, Morioka S, Ohta S, Ueki J;

XX WPI; 1998-086573/08.

XX Method for expressing foreign gene - comprises inserting gene
PT downstream from promoter comprising at least two introns, useful in,
PT e.g. genetic engineering

XX Claim 4; Page 10; 20pp; Japanese.

XX This sequence represents an intron from the Oryza sativa ubiquitin gene,
CC and can be used in the method of the invention. The method is for
CC expressing a foreign gene by inserting the gene downstream from a
CC promoter, comprising at least 2 intron sequences. The method is used for
CC expression of foreign genes, useful in, e.g. genetic engineering. The
CC method achieves higher levels of expression than conventional methods.

XX Sequence 173 BP; 21 A; 48 C; 47 G; 57 T; 0 other;

Query Match 100.0%; Score 173; DB 19; Length 173;
Best Local Similarity 100.0%; Pred. No. 8.8e-45;
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaagcccaagtgtgcttagagctaaagcgcactagagctcttgcgtctctctcc 60

DB 1 gtaagcccaagtgtgcttagagctaaagcgcactagagctcttgcgtctctctcc 60
|||||

QY 61 gctcagatctgctgtgctgtgcttgcgtcagaacctactctgtgcgagtgctgct 120
|||||
DB 61 gctcagatctgctgtgctgtgcttgcgtcagaacctactctgtgcgagtgctgct 120

QY 121 gctcgtctctctcctcccaagttcgatctgctgtgtggtggtggtggtggtggt 173
|||||
DB 121 gctcgtctctctcctcccaagttcgatctgctgtgtggtggtggtggtggtggt 173

RESULT 3

T42851 standard; DNA; 183 BP.

XX T42851;

XX 16-JUN-1997 (first entry)

XX Promoter #1 for phospholipase D gene.

XX Phospholipase D; rice; promoter; ss.

XX Oryza sativa.

XX WO9630510-A1.

XX 03-OCT-1996.

XX 28-MAR-1996; 96WO-JP00812.

XX 29-MAR-1995; 95JP-0096126.

XX (NISR) JAPAN TOBACCO INC.

XX Morioka S, Ueki J;

XX WPI; 1996-455357/45.

XX Promoter DNA sequence derived from rice - used to increase
PT expression of foreign genes in transformed hosts

XX Claim 1; Page 14; 29pp; Japanese.

XX T42851 and T42852 represent promoters isolated from rice. These sequences
CC are specifically promoters for the phospholipase D gene (PLD). T42852
CC represents a shortened version of this sequence. These sequences are
CC efficient promoters for greatly increasing the expression of foreign
CC genes in transformant rice and other plants.

XX Sequence 183 BP; 23 A; 51 C; 51 G; 58 T; 0 other;

Query Match 100.0%; Score 173; DB 17; Length 183;
Best Local Similarity 100.0%; Pred. No. 9e-45;
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaagcccaagtgtgcttagagctaaagcgcactagagctcttgcgtctctctcc 60
|||||
DB 6 gtaagcccaagtgtgcttagagctaaagcgcactagagctcttgcgtctctctcc 65
61 gctcagatctgctgtgctgtgcttgcgtcagaacctactctgtgcgagtgctgct 120
61 gctcagatctgctgtgctgtgcttgcgtcagaacctactctgtgcgagtgctgct 120
DB 66 gctcagatctgctgtgctgtgcttgcgtcagaacctactctgtgcgagtgctgct 125
66 gctcagatctgctgtgctgtgcttgcgtcagaacctactctgtgcgagtgctgct 125
QY 121 gctcgtctctctcctcccaagttcgatctgctgtgtggtggtggtggtggtggt 173
|||||
DB 126 gctcgtctctctcctcccaagttcgatctgctgtgtggtggtggtggtggtggt 178

RESULT 4

QY 086785

[illegible]

XX		Phospholipase D gene sequence fragment.
DE		
XX		
KW		Phospholipase D; rice; promoter; ss.
XX		
OS	Oryza sativa.	
XX		
FH	Key	Location/Qualifiers
FT	exon	1876..1983
FT		/tag= a
FT		/number= 1
FT	Intron	/note= "encodes residues 1 to 36 of W06134"
FT		1984..2523
FT		/tag= b
FT	exon	/number= 1
FT		2524..2799
FT		/tag= c
FT		/number= 2
FT		/note= "encodes residues 37 to 128 of W06134"
PN		
PD	W09630510-A1.	
PD		
PD	03-OCT-1996.	
PF		
PF	28-MAR-1996;	96WO-JP00812.
PR		
PR	29-MAR-1995;	95JP-0096126.
PA	(NISB) JAPAN TOBACCO INC.	
PI	Morioka S, Ueki J;	
PI		
DR	WPI: 1996-455357/45.	
DR	P-PSDB; W06134.	
PT		
PT	Promoter DNA sequence derived from rice - used to increase expression of foreign genes in transformed hosts	
PS		
PS	Disclosure; Page 20-22; 29pp; Japanese.	
CC		
CC	This sequence represents a fragment of the coding sequence of the rice phospholipase D gene (PLD). The promoter for the PLD gene was isolated using the primers shown in T42857 and T42588. The promoters (see T42851 and T42852) are efficient promoters for greatly increasing the expression of foreign genes in transformant rice and other plants.	
CC		
CC		
SQ	Sequence 2799 BP; 692 A; 709 C; 609 G; 789 T; 0 other:	
	Query Match	100.0%; Score 173; DB 17; Length 2799;
	Best Local Similarity	100.0%; Pred. No. 1.9e-44;
	Matches 173; Conservative	0; Mismatches 0; Indels 0; Gaps 0
QY	1 gtaagccaagtgtgcttgaagctaagcgcaactagaagcttcgtcgcgtgtctctcc 60 	
Db	1666 gtaagcccagttgtcttaagctaagcgcaactagaagcttcgtcgcgttcctcc 1725	
QY	61 gtccgatcttgcttgcttgcttgcttgcttgctgaacctactctgtctcgagatgtcgt 120 	
Db	1726 gtccgatcttgcttgcttgcttgcttgctgaacctactctgtctcgagatgtcgt 1785	
QY	121 gtctgctctctccctcaagctcgatctgtgtgtggggggcgacaa 173 	
Db	1786 gtctgctctctctccctcaagctcgatctgtgtgtgtggggggcgacaa 1838	
RESULT	6	
ID	XI3856	
XX	XI3856 standard; DNA; 796 BP.	
AC	XI3856;	
XX		
DT	19-MAR-1999 (first entry)	

XX		Enterococcus faecalis genome contig SEQ ID NO:919.
DE		
XX		
KW	Enterococcus faecalis; contig; detection; Enterococcal infection;	
KW	vaccine; attenuation; computer readable medium; ds.	
XX		
OS	Enterococcus faecalis.	
PN	WO9850555-A2.	
XX		
PD	12-NOV-1998.	
XX		
PE	04-MAY-1998; 98WO-US08985.	
PR	14-NOV-1997; 97US-0066009.	
PR	06-MAY-1997; 97US-0044031.	
XX	16-MAY-1997; 97US-0046655.	
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI		
DR	Barash SC, Dillon PJ, Kunsch CA:	
XX	WPI: 1999-045171/04.	
PT	New isolated Enterococcus faecalis polynucleotides and polypeptides	
PT	- used to develop products for the detection of Enterococcus and for	
PT	use in vaccines for prevention or attenuation of Enterococcus	
PT	Infection.	
PS	Claim 1; Page 2047-2048; 2084bp; English.	
XX		
CC	A computer readable medium has been developed which has recorded on it	
CC	982 nucleotide sequences isolated from the Enterococcus faecalis genome.	
CC	X12938 to X13919 represent these nucleotide sequences which are primary	
CC	nucleotide sequences, also known as contigs. The computer-based system	
CC	can identify fragments of the Enterococcus faecalis genome with	
CC	commercial importance. The products can be used to detect the presence	
CC	of Enterococcus faecalis in samples. They can also be used for	
CC	diagnosing Enterococcal infection in an animal and monitoring	
CC	progression of disease, and for identifying agents which can be used to	
CC	modulate the growth or pathogenicity of Enterococcus faecalis, or	
CC	another related organism, in vivo or in vitro. In particular the	
CC	polypeptides encoded by the Enterococcus faecalis nucleotide sequences	
CC	can be used in vaccines to prevent or attenuate an Enterococcal	
CC	infection.	
XX		
SQ	Sequence 796 BP; 113 A; 170 C; 163 G; 346 T; 4 other;	
OY	Query Match	19.2%; Score 33.2; DB 20; Length 796;
	Best Local Similarity	53.0%; Pred. No. 0.21;
	Matches	71; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
DB	11 tgcgttagcgaacgacagctagatcttcgtcgctgctcttcgcgtcacagtct 70	
DB	385 ttgtcttcgtcgctgctcacactctcgtgctgcgtgtgttgatctgccacagtttgt 444	
OY	71 gcttgcctgctgcttcgcctagaccctacctcgtcgtcgcagatgctgcgttcctt 130	
DB	445 gcttcttcaatgatcgtcgtgtagacttcacgctcgcgtcggtgtagtattcttcctg 504	
OY	131 ccttcctcaagtcc 144	
DB	505 tgattctcgttttc 518	
RESULT	7	
X20580		
ID	X20580 standard; DNA; 19142 BP.	
XX		
AC	X20580;	
XX		
DT	05-MAY-1999 (first entry)	

[illegible]

OY 115 gtgcgtctcgtctctctcctcctcaagtt 143
 Db 1181 CCCCACAGTAGTGTCTGCTTCATGTT 1153

RESULT 13

X20167/c
 ID X20167 standard; DNA: 586 BP.

XX X20167;

DT 20-APR-1999 (first entry)

DE Enterococcus faecalis EF088 gene fragment.

KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 detection; attenuation; antigenic; ss.

OS Enterococcus faecalis.

PN W09850554-A2.

PD 12-NOV-1998.

PE 04-MAY-1998; 98WO-US08959.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Bailey C, Choi GH, Hromocky J A, Kunsch CA;

DR WPI: 1999-070095/06.

DR P-PSDB; Y00177.

PT New isolated Enterococcus faecalis polynucleotides - used to develop
 PT products for the detection of Enterococcus and for use in vaccines
 PT for prevention or attenuation of Enterococcus infection

PS Claim 1; Page 181; 301pp; English.

CC The present sequence encodes an antigenic polypeptide fragment
 CC isolated from Enterococcus faecalis. The present invention describes
 CC genes, proteins and antigenic polypeptides isolated from E. faecalis.
 CC The proteins can be used in vaccines for preventing or attenuating an
 CC infection caused by a member of the Enterococcus genus in an animal.
 CC They can also be used for detecting Enterococcus antibodies in a sample.
 CC The nucleotide sequences can be used for detecting Enterococcus nucle-
 CC acids. Products from the present invention can also be used for
 CC screening compounds to identify agonists and antagonists of E. faecalis
 CC protein activity.

SO Sequence 586 BP; 263 A; 106 C; 131 G; 86 T; 0 other;

Query Match 16.6%; Score 28.8; DB 20; Length 586;
 Best Local Similarity 50.7%; Pred. No. 4.4;

Matches 69; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 9 agtgcgttaagcctaagcgaactctctgctgcgtctctctctcgcgcagat 68
 Db 179 ACTGCTTTCTGCTGCTACTTCTGCTGCTGCTGCTTTGATCTGCCACAGTTT 120

OY 69 ctgcgtctgctgcgtgcctaagaccctactctgctgcgagatgctgcgtctgc 128
 Db 119 GTGCTTGTTCATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60

OY 129 ttcctctcctcaagtc 144
 Db 59 GGTGTTCTTTTGTTC 44

RESULT 14
 X20162/c
 ID X20162 standard; DNA: 2559 BP.

XX X20162;

DT 20-APR-1999 (first entry)

DE Enterococcus faecalis gene EF086.

KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 detection; attenuation; antigenic; ss.

OS Enterococcus faecalis.

PN W09850554-A2.

PD 12-NOV-1998.

PE 04-MAY-1998; 98WO-US08959.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Bailey C, Choi GH, Hromocky J A, Kunsch CA;

DR WPI: 1999-070095/06.

DR P-PSDB; Y00172.

PT New isolated Enterococcus faecalis polynucleotides - used to develop
 PT products for the detection of Enterococcus and for use in vaccines
 PT for prevention or attenuation of Enterococcus infection

PS Claim 1; Page 176-177; 301pp; English.

CC The present sequence represents a gene isolated from
 CC Enterococcus faecalis. The present invention describes genes, proteins
 CC and antigenic polypeptides isolated from E. faecalis. The proteins can
 CC be used in vaccines for preventing or attenuating an infection caused
 CC by a member of the Enterococcus genus in an animal. They can also be
 CC used for detecting Enterococcus antibodies in a sample. The nucleotide
 CC sequences can be used for detecting Enterococcus nucleic acids.
 CC Products from the present invention can also be used for screening
 CC compounds to identify agonists and antagonists of E. faecalis protein
 CC activity.

SO Sequence 2559 BP; 923 A; 427 C; 539 G; 668 T; 2 other;

Query Match 16.6%; Score 28.8; DB 20; Length 2559;
 Best Local Similarity 50.7%; Pred. No. 6.6;

Matches 69; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 9 agtgcgttaagcctaagcgaactctctgctgcgtctctctcgcgcagat 68
 Db 2068 ACTGCTTTCTGCTGCTACTTCTGCTGCTGCTGCTTTGATCTGCCACAGTTT 2009

OY 69 ctgcgtctgctgcgtgcctaagaccctactctgctgcgagatgctgcgtctgc 128
 Db 2008 GTGCTTGTTCATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1949

OY 129 ttcctctcctcaagtc 144
 Db 1948 GGTGTTCTTTTGTTC 1933

RESULT 15

X20164/c

ID	X20164 standard; DNA: 2559 BP.
AC	X20164;
DY	20-APR-1999 (first entry)
DE	Enterococcus faecalis gene EF087.
KW	Enterococcus faecalis; infection; vaccine; immune response; diagnosis; detection; attenuation; antigenic; ss.
OS	Enterococcus faecalis.
PN	M09850554-A2.
PD	12-NOV-1998.
PF	04-MAY-1998; 98WO-US08959.
PR	14-NOV-1997; 97US-0066009. 06-MAY-1997; 97US-0044031.
PR	16-MAY-1997; 97US-0046655.
PA	(HUMA-) HUMAN GENOME SCI INC.
PI	Bailey C, Choi GH, Hiromocky} A, Kunsch CA;
DR	WPI: 1999-070095/06. P-PSDB: Y00174.
PT	New Isolated Enterococcus faecalis polynucleotides - used to develop products for the detection of Enterococcus and for use in vaccines for prevention or attenuation of Enterococcus infection
PS	Claim 1; Page 178-179; 301pp; English.
CC	The present sequence represents a gene isolated from Enterococcus faecalis. The present invention describes genes, proteins and antigenic polypeptides isolated from E. faecalis. The proteins can be used in vaccines for preventing or attenuating an infection caused by a member of the Enterococcus genus in an animal. They can also be used for detecting Enterococcus antibodies in a sample. The nucleotide sequences can be used for detecting Enterococcus nucleic acids. Products from the present invention can also be used for screening compounds to identify agonists and antagonists of E. faecalis protein activity.
SQ	Sequence 2559 BP; 923 A; 427 C; 539 G; 668 T; 2 other:
Query Match	16.6%; Score 28.8; DB 20; Length 2559;
Best Local Similarity	50.7%; Fred. No. 6.6;
Matches 69; Conservative	0; Mismatches 67; Indels 0; Gaps 0;
OY	9 agtgcgttaagcgaacgactagaagcttcgtcgcttgctctctccgcgcagat 68
Db	2068 ACCTGCTTTCGCTTGCGCTACTTCTGCTGGGTGCTTTTGATCAGTGCACAGTT 2009
OY	69 ctgcgttcgttcgttcgttcgttagaacctactctgctgcgagtgctgcgtcgc 128
Db	2008 GTGCTTGTTCGAATGCTTCGCTGTGAGACTTCCACACTGCTTCGGGTGAGTCTTTTCTT 1949
OY	129 ttctctcctcaagtcc 144
Db	1948 GGTCCTTCCTTTTGTTC 1933

Search completed: February 12, 2001, 21:28:01
Job time: 3557 sec

1	NAME/KEY: misc_feature	
2	LOCATION: complement (937..2262)	
3	OTHER INFORMATION: /label= ORF1SUBSEQUENCE	
4	FEATURE:	
5	NAME/KEY: misc_feature	
6	LOCATION: complement (2649..3458)	
7	OTHER INFORMATION: /label= ORF2SUBSEQUENCE	
8	FEATURE:	
9	NAME/KEY: misc_feature	
10	LOCATION: 3726..4799	
11	OTHER INFORMATION: /label= ORF3SUBSEQUENCE	
12	FEATURE:	
13	NAME/KEY: misc_feature	
14	LOCATION: complement (4607..4918)	
15	OTHER INFORMATION: /label= ORF5SUBSEQUENCE	
16	FEATURE:	
17	NAME/KEY: misc_feature	
18	LOCATION: 5143..6216	
19	OTHER INFORMATION: /label= ORF6SUBSEQUENCE	
20	FEATURE:	
21	NAME/KEY: misc_feature	
22	LOCATION: complement (5071..5643)	
23	OTHER INFORMATION: /label= ORF7SUBSEQUENCE	
24	FEATURE:	
25	NAME/KEY: misc_feature	
26	LOCATION: 6609..8888	
27	OTHER INFORMATION: /label= ORF8SUBSEQUENCE	
28	FEATURE:	
29	NAME/KEY: misc_feature	
30	LOCATION: complement (6576..6830)	
31	OTHER INFORMATION: /label= ORF9SUBSEQUENCE	
32	FEATURE:	
33	NAME/KEY: misc_feature	
34	LOCATION: 9748..10044	
35	OTHER INFORMATION: /label= ORF10SUBSEQUENCE	
36	FEATURE:	
37	NAME/KEY: misc_feature	
38	LOCATION: complement (10509..11282)	
39	OTHER INFORMATION: /label= ORF11SUBSEQUENCE	
40	FEATURE:	
41	NAME/KEY: misc_feature	
42	LOCATION: 12466..13002	
43	OTHER INFORMATION: /label= ORF12SUBSEQUENCE	
44	FEATURE:	
45	NAME/KEY: misc_feature	
46	LOCATION: 13723..14319	
47	OTHER INFORMATION: /label= ORF13SUBSEQUENCE	
48	FEATURE:	
49	NAME/KEY: misc_feature	
50	LOCATION: 15659..16210	
51	OTHER INFORMATION: /label= ORF14SUBSEQUENCE	
52	FEATURE:	
53	NAME/KEY: misc_feature	
54	LOCATION: complement (16517..17545)	
55	OTHER INFORMATION: /label= ORF15SUBSEQUENCE	
56	FEATURE:	
57	NAME/KEY: misc_feature	
58	LOCATION: complement (17737..18189)	
59	OTHER INFORMATION: /label= ORF16SUBSEQUENCE	
60	FEATURE:	
61	NAME/KEY: misc_feature	
62	LOCATION: complement (18177..18743)	
63	OTHER INFORMATION: /label= ORF17SUBSEQUENCE	
64	FEATURE:	
65	NAME/KEY: misc_feature	
66	LOCATION: complement (19031..19390)	
67	OTHER INFORMATION: /label= ORF18SUBSEQUENCE	
68	NAME/KEY: misc_feature	
69	LOCATION: complement (19031..19390)	
70	OTHER INFORMATION: /label= ORF18SUBSEQUENCE	
71	NAME/KEY: misc_feature	
72	LOCATION: complement (19031..19390)	
73	OTHER INFORMATION: /label= ORF18SUBSEQUENCE	
74	NAME/KEY: misc_feature	
75	LOCATION: complement (19031..19390)	
76	OTHER INFORMATION: /label= ORF18SUBSEQUENCE	
77	NAME/KEY: misc_feature	
78	LOCATION: complement (19031..19390)	
79	OTHER INFORMATION: /label= ORF18SUBSEQUENCE	
80	NAME/KEY: misc_feature	
81	LOCATION: complement (19031..19390)	
82	OTHER INFORMATION: /label= ORF18SUBSEQUENCE	
83	NAME/KEY: misc_feature	
84	LOCATION: complement (19031..19390)	
85	OTHER INFORMATION: /label= ORF18SUBSEQUENCE	
86	NAME/KEY: misc_feature	
87	LOCATION: complement (19031..19390)	
88	OTHER INFORMATION: /label= ORF18SUBSEQUENCE	
89	NAME/KEY: misc_feature	
90	LOCATION: complement (19031..19390)	
91	OTHER INFORMATION: /label= ORF18SUBSEQUENCE	
92	NAME/KEY: misc_feature	
93	LOCATION: complement (19031..19390)	
94	OTHER INFORMATION: /label= ORF18SUBSEQUENCE	
95	NAME/KEY: misc_feature	
96	LOCATION: complement (19031..19390)	
97	OTHER INFORMATION: /label= ORF18SUBSEQUENCE	
98	NAME/KEY: misc_feature	
99	LOCATION: complement (19031..19390)	
100	OTHER INFORMATION: /label= ORF18SUBSEQUENCE	

[illegible]

```

1      RESULT 13
2      US-08-459-569-19
3      Sequence 19, Application US/08459569
4      Patent No. 5543501
5      GENERAL INFORMATION:
6      APPLICANT: Slightom, Jerry L.
7      APPLICANT: Tepfer, David A.
8      TITLE OF INVENTION: R1 T-DNA Promoters
9      NUMBER OF SEQUENCES: 19
10     CORRESPONDENCE ADDRESS:
11     ADDRESSEE: GRAY, CARV, AMES & FRYE
12     STREET: 401 B Street, Suite 1700
13     CITY: San Diego
14     STATE: California
15     COUNTRY: USA
16     ZIP: 92101-4297
17
18     COMPUTER READABLE FORM:
19     MEDIUM TYPE: Floppy disk
20     COMPUTER: IBM PC compatible
21     OPERATING SYSTEM: PC-DOS/MS-DOS
22     SOFTWARE: PatentIn Release #1.0, Version #1.25
23     CURRENT APPLICATION DATA:
24     APPLICATION NUMBER: US/08/459,569
25     FILING DATE: 02-JUN-1995
26     CLASSIFICATION: 530
27     PRIOR APPLICATION DATA:
28     APPLICATION NUMBER: US 08/008,216
29     FILING DATE: 25-JAN-1993
30     APPLICATION NUMBER: US 06/7725,368
31     FILING DATE: 22-APR-1985
32     ATTORNEY/AGENT INFORMATION:
33     NAME: Bairdhorst, Marine W.
34     REGISTRATION NUMBER: 36,740
35     REFERENCE/DOCKET NUMBER: P1020051
36     TELECOMMUNICATION INFORMATION:
37     TELEPHONE: (619) 699-2700
38     TELEFAX: (619) 236-1048
39     INFORMATION FOR SEQ ID NO: 19:
40     SEQUENCE CHARACTERISTICS:
41     LENGTH: 21126 base pairs
42     TYPE: nucleic acid
43     STRANDEDNESS: single
44     TOPOLOGY: linear
45     MOLECULE TYPE: DNA (genomic)
46     HYPOTHETICAL: NO
47     ORIGINAL SOURCE:
48     ORGANISM: Agrobacterium rhizogenes
49     STRAIN: STRAIN A4
50     IMMEDIATE SOURCE:
51     LIBRARY: CONVOLVULUS ARVENSIS PLANT CELLS
52     CLONE: CLONE 7
53
54     FEATURE:
55     NAME/KEY: misc_feature
56     LOCATION: complement (937..2262)
57     OTHER INFORMATION: /label=ORF1SUBSEQUENCE
58     FEATURE:
59     NAME/KEY: misc_feature
60     LOCATION: complement (2649..3458)
61     OTHER INFORMATION: /label=ORF2SUBSEQUENCE

```

```

1 FEATURE:
2 NAME/KEY: misc_feature
3 LOCATION: 3726..4799
4 OTHER INFORMATION: /label= ORF3SUBSEQUENCE
5
6 FEATURE:
7 NAME/KEY: misc_feature
8 LOCATION: complement (4041..4400)
9 OTHER INFORMATION: /label= ORF4SUBSEQUENCE
10
11 FEATURE:
12 NAME/KEY: misc_feature
13 LOCATION: complement (4607..4918)
14 OTHER INFORMATION: /label= ORF5SUBSEQUENCE
15
16 FEATURE:
17 NAME/KEY: misc_feature
18 LOCATION: 5143..6216
19 OTHER INFORMATION: /label= ORF6SUBSEQUENCE
20
21 FEATURE:
22 NAME/KEY: misc_feature
23 LOCATION: complement (5071..5643)
24 OTHER INFORMATION: /label= ORF7SUBSEQUENCE
25
26 FEATURE:
27 NAME/KEY: misc_feature
28 LOCATION: 6609..8888
29 OTHER INFORMATION: /label= ORF8SUBSEQUENCE
30
31 FEATURE:
32 NAME/KEY: misc_feature
33 LOCATION: complement (6576..6830)
34 OTHER INFORMATION: /label= ORF9SUBSEQUENCE
35
36 FEATURE:
37 NAME/KEY: misc_feature
38 LOCATION: 9748..10044
39 OTHER INFORMATION: /label= ORF10SUBSEQUENCE
40
41 FEATURE:
42 NAME/KEY: misc_feature
43 LOCATION: complement (10509..11282)
44 OTHER INFORMATION: /label= ORF11SUBSEQUENCE
45
46 FEATURE:
47 NAME/KEY: misc_feature
48 LOCATION: 12466..13002
49 OTHER INFORMATION: /label= ORF12SUBSEQUENCE
50
51 FEATURE:
52 NAME/KEY: misc_feature
53 LOCATION: 13723..14319
54 OTHER INFORMATION: /label= ORF13SUBSEQUENCE
55
56 FEATURE:
57 NAME/KEY: misc_feature
58 LOCATION: 15659..16210
59 OTHER INFORMATION: /label= ORF14SUBSEQUENCE
60
61 FEATURE:
62 NAME/KEY: misc_feature
63 LOCATION: complement (16517..17545)
64 OTHER INFORMATION: /label= ORF15SUBSEQUENCE
65
66 FEATURE:
67 NAME/KEY: misc_feature
68 LOCATION: complement (17737..18189)
69 OTHER INFORMATION: /label= ORF16SUBSEQUENCE
70
71 FEATURE:
72 NAME/KEY: misc_feature
73 LOCATION: complement (18177..18743)
74 OTHER INFORMATION: /label= ORF17SUBSEQUENCE
75
76 FEATURE:
77 NAME/KEY: misc_feature
78 LOCATION: complement (19031..19390)
79 OTHER INFORMATION: /label= ORF18SUBSEQUENCE
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
9
```

[illegible][illegible]

```

RESULT 14
US-08-458-031-19
Sequence 19, Application US/08458031
Patent No. 5824866
GENERAL INFORMATION:
APPLICANT: Slightom, Jerry L.
APPLICANT: Tepfer, David A.
TITLE OF INVENTION: RI T-DNA Promoters
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: GRAY, CARY, AMES & FRYE
STREET: 401 B Street, Suite 1700
City: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-4297
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,831
FILING DATE: 02-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/008,216
FILING DATE: 25-JAN-1993
APPLICATION NUMBER: US 06/725,368
FILING DATE: 22-APR-1985
ATTORNEY/AGENT INFORMATION:
NAME: Barthorst, Marile W.
REGISTRATION NUMBER: 36,740
REFERENCE/DOCKET NUMBER: P1020U51
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 699-2700
TELEFAX: (619) 236-1048
INFORMATION FOR SEQ. ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 21126 base pairs
type: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Agrobacterium rhizogenes
STRAIN: STRAIN A4
IMMEDIATE SOURCE:
LIBRARY: CONVOLUTULUS ARVENSIS PLANT CELLS
CLONE: CLONE 7
FEATURE:
NAME/KEY: misc_feature
LOCATION: complement (937..2262)
OTHER INFORMATION: /label=ORF1SUBSEQUENCE
FEATURE:
NAME/KEY: misc_feature
LOCATION: complement (2649..3458)
OTHER INFORMATION: /label=ORF2SUBSEQUENCE
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3726..4799
OTHER INFORMATION: /label=ORF3SUBSEQUENCE
FEATURE:
NAME/KEY: misc_feature
LOCATION: complement (4041..4400)

```



```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	Similarity	Score	DB 1	Length
Best Local	52.8%	Pred. No. 23		21126
Matches	57	Conservative	0	Mismatches 51
				Indels 0
				Gaps 0

Qy	44	tcgtctgctctcttcgcgcgtcagatctgcttgccttgcttgcttgctgctagataacctactct	103
Db	3673	TTGCTGGCCACCATTTTCTTAAATCTGCCCCCTTTCGGCCCTCGGTATCATGCCGCCCA	3732
Qy	104	gtctctgagagtgtcgcctgcttcgtcttccttcctcaagtcgatactga	151
Db	3733	ACGAGCGCCCATCGCTCGCCCGGCGATTTGCCCTCCCGAGGCGCAACAGGA	3780

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	15.1%	Score 26.2;	DB 1;	Length 6453;
Best Local Similarity	53.4%;	Pred. No. 18;		
Matches	55;	Conservative	48;	Indels 0; Gaps 0;
QY	70	tgtctgcttgcttgcttcgtcagaaacctacatctgtgctgcgaagtgcgcgttgcgtct	129	
Db	3552	ttcccttccttccttccttccttccttccttccttccttccttccttccttccttccttccttc	3493	
QY	130	tccttcctcaagtcgatcatctgattgtgtggtggaggcgaca	172	
Db	3492	atccggcacctccatgcctctgcagcttgctggggggggcacaca	3450	

Search completed: February 12, 2001, 21:25:33
Job time: 3419 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2001, 21:02:16 ; Search time 1923.53 Seconds
(without alignments)
630.245 Million cell updates/sec

Title: US-09-600-602-3

Perfect score: 173
Sequence: 1 gtaagccagtggtgcttagg.....gtgtgtgtggtggggcgccag 173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: em_estba.*
42: em_estfun.*
43: em_esthum1.*

44: em_esthum2.*
45: em_esthum3.*
46: em_esthum4.*
47: em_esthum5.*
48: em_esthum6.*
49: em_esthum7.*
50: em_esthum8.*
51: em_esthum9.*
52: em_esthum10.*
53: em_esthum11.*
54: em_esthum12.*
55: em_esthum13.*
56: em_esthum14.*
57: em_esthum15.*
58: em_esthum16.*
59: em_esthum17.*
60: em_esthum18.*
61: em_esthum19.*
62: em_esthum20.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estov1.*
68: em_estov2.*
69: em_estp11.*
70: em_estp12.*
71: em_estp13.*
72: em_estp14.*
73: em_estp15.*
74: em_estro1.*
75: em_estro2.*
76: em_estro3.*
77: em_estro4.*
78: em_estro5.*
79: em_estro6.*
80: em_estro7.*
81: em_estro8.*
82: em_estro9.*
83: em_estro10.*
84: em_estro11.*
85: em_estro12.*
86: em_estro13.*
87: gb_esta1.*
88: gb_esta2.*
89: gb_esta3.*
90: gb_esta4.*
91: gb_esta5.*
92: gb_esta6.*
93: gb_esta7.*
94: gb_esta8.*
95: gb_esta9.*
96: gb_est10.*
97: gb_est11.*
98: gb_est12.*
99: gb_est13.*
100: gb_est14.*
101: gb_est15.*
102: gb_est16.*
103: gb_est17.*
104: gb_est18.*
105: gb_est19.*
106: gb_est20.*
107: gb_est21.*
108: gb_est22.*
109: gb_est23.*
110: gb_est24.*
111: gb_est25.*
112: em_esthum21.*
113: em_esthum22.*
114: em_esthum23.*
115: em_estom1.*
116: em_estom2.*

117: em_estp16:*
 118: em_estp17:*
 119: em_estp18:*
 120: em_estp14:*
 121: em_estp15:*
 122: em_estp16:*
 123: em_estp17:*
 124: em_estp18:*
 125: em_estp19:*
 126: gb_est58:*
 127: gb_est59:*
 128: gb_est60:*
 129: gb_est61:*
 130: gb_est62:*
 131: gb_est63:*
 132: gb_est64:*
 133: gb_est65:*
 134: gb_est66:*
 135: gb_est67:*
 136: gb_est68:*
 137: gb_est69:*
 138: gb_est70:*
 139: gb_est71:*
 140: gb_est72:*
 141: gb_est73:*
 142: gb_est74:*
 143: gb_est75:*
 144: gb_est76:*
 145: gb_est77:*
 146: gb_est78:*
 147: gb_est79:*
 148: gb_est80:*
 149: gb_est81:*
 150: gb_est82:*
 151: gb_est83:*
 152: gb_est84:*
 153: gb_est85:*
 154: gb_est86:*
 155: gb_est87:*
 156: gb_est88:*
 157: gb_est89:*
 158: gb_est90:*
 159: gb_est91:*
 160: gb_est92:*
 161: gb_est93:*
 162: gb_est94:*
 163: gb_est95:*
 164: gb_est96:*
 165: gb_est97:*
 166: gb_est98:*
 167: gb_est99:*
 168: gb_est100:*
 169: gb_est101:*
 170: gb_est102:*
 171: gb_est103:*
 172: gb_est104:*
 173: gb_est105:*
 174: gb_est106:*
 175: gb_est107:*
 176: gb_est108:*
 177: gb_est109:*
 178: gb_est110:*
 179: gb_est111:*
 180: gb_est112:*
 181: gb_est113:*
 182: gb_est114:*
 183: gb_est115:*
 184: gb_est116:*
 185: gb_est117:*
 186: gb_est118:*
 187: gb_est119:*
 188: gb_est120:*
 189: gb_est121:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	37.2	21.5	668	94	AM694929	AM694929 NF081E10S
2	37.2	21.5	671	94	AM694702	AM694702 NF079B05S
3	36.8	21.3	733	139	BF036458	BF036458 601460209
4	36.2	20.9	1078	190	CNS0074P	AL075925 Drosophila
5	35.8	20.7	974	137	BE880847	BE880847 601493284
6	35	20.2	743	109	BE519618	BE519618 HV_CED001
7	34.4	19.9	632	97	AM974965	AM974965 EST387070
8	34.2	19.8	990	193	CNS04YVF	AL313476 Tetradon
9	33.8	19.5	702	106	BE268416	BE268416 601124712
10	33.8	19.5	915	135	BE798838	BE798838 601584668
11	33.6	19.4	659	182	A2463907	A2463907 1M0273K03
12	33.6	19.4	997	190	CNS005TE	AL121302 uc30b03.r
13	33.4	19.3	443	16	AI121302	AI121302 uc30b03.r
14	33.4	19.3	584	27	AI981330	AI981330 pat. PK005
15	33.4	19.3	1265	139	BF032829	BF032829 601455658
16	32.8	19.0	473	10	MA703492	MA703492 z11h02.s
17	32.8	19.0	618	28	AJ398711	AJ398711 AJ398711
18	32.8	19.0	706	19	AI401719	AI401719 th30b10.x
19	32.8	19.0	821	176	A2202926	A2202926 SP_0079_B
20	32.8	19.0	876	192	CNS041KE	AL270311 Tetradon
21	32.8	19.0	893	105	BE259350	BE259350 601106438
22	32.8	19.0	900	105	BE249907	BE249907 600942952
23	32.6	18.8	693	160	AO656370	AO656370 Sheared D
24	32.6	18.8	746	107	BE415895	BE415895 MUG002.B0
25	32.4	18.7	903	94	AW728116	AW728116 GA_Ea001
26	32.4	18.7	1093	192	CNS03RTE	AL257676 Tetradon
27	32	18.5	484	92	AM577552	AM577552 PM2-B7054
28	32	18.5	876	175	A2138727	A2138727 SP_0170_A
29	32	18.5	998	135	BE799289	BE799289 601591751
30	32	18.5	1101	190	CNS017GS	AL108022 Drosophila
31	32	18.5	1201	190	CNS0167A	AL106384 Drosophila
32	32	18.5	1203	190	CNS015VR	AL106077 Drosophila
33	31.8	18.4	846	190	CNS010RJ	AL099337 Drosophila
34	31.8	18.4	897	138	BE973857	BE973857 601680665
35	31.6	18.3	356	193	FR0029482	AL025851 Fugu tubr
36	31.6	18.3	479	37	AV635060	AV635060 AV635060
37	31.6	18.3	620	180	A2358273	A2358273 1M0100B11
38	31.6	18.3	648	110	BE593348	BE593348 WSI_100_F
39	31.6	18.3	814	109	BE563541	BE563541 601334834
40	31.6	18.3	972	135	BE740157	BE740157 601595106
41	31.6	18.3	1235	137	BE881061	BE881061 601492122
42	31.6	18.3	1627	137	BE905523	BE905523 601495154
43	31.6	18.3	1822	138	BE967064	BE967064 601660603
44	31.4	18.2	353	146	W00065	W00065 T9EST74F0
45	31.4	18.2	360	39	AM086231	AM086231 xc70C10.x

ALIGNMENTS

RESULT 1
 LOCUS AM694929 668 bp mRNA
 DEFINITION NF081E10ST11082 Developing stem Medicago truncatula cDNA clone
 ACCESSION AM694929
 VERSION AM694929.1 GI:7569691
 KEYWORDS EST.
 SOURCE barrel medic.

	ORGANISM	Medicago truncatula Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; core eucots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago. 1 (bases 1 to 68)					
REFERENCE	AUTHORS	He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell, ,C.J., Flores,H.R., Imman,J.T., Weller,J.W., May,G.D. and Dixon, ,R.A.					
TITLE	JOURNAL COMMENT	<p>Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula stem library Unpublished (2000) Contact: Dixon RA Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73402, USA Tel: 580 221 7302 Fax: 580 221 7380 Email: radixon@noble.org Insert Length: 668 Std Error: 0.00 Plate: 081 row: E column: 10 Seq primer: TCACACAGAAACAAGCATGTGAC. Location/Qualifiers 1..668 /organism="Medicago truncatula" /db_xref="taxon:3880" /cclone="NF081E10ST" /cloned_lib="Developing stem" /tissue_type="stem" /dev_stage="Pooled developmental" /note="Vector: Lambda Zap; Contains a mixture of internodal stem segments"</p>					
FEATURES	SOURCE						
BASE COUNT	ORIGIN	330 a 62 c 188 g 88 t					
Query Match:	Best Local Similarity	21.5%; Score 37.2; DB 94; Length 668;					
Matches	Conservative	59.4%; Pred.No.1.1; Mismatches 43; Indels 0; Gaps 0;					
OY	Db	48	tgtcttcttcacgcgtagcatcgttgttgctggctcgtaagactccctctgtgc	107			
I	I			I	I	I	I
Db	432	TTTCCTCCATCGTGCCTGGTGCTGCCATTCACTCTTTACTTAATACTAATTCGT	373				
OY	108	tgcgaagtgcgctgccctctcctcccctaagtcgatctgatt	153				
I	I		I	I	I	I	I
Db	372	TTCACTTGTTCTTTCATCTCTCTTGCTTACTCTTTTCCTCGGAATT	327				
RESULT	2						
Locus	AM694702/C	671 bp mRNA EST	15-JUN-2000				
DEFINITION	NF079B05ST 5'	Developing stem Medicago truncatula cDNA clone					
ACCSSION	NF079B05ST						
VERSION	AW694702						
KEYWORDS	AM694702.1 GI:7569464						
SOURCE	EST						
ORGANISM	barell medic.						
	Medicago truncatula						
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; core eucots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago. 1 (bases 1 to 671) He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell, ,C.J., Flores,H.R., Imman,J.T., Weller,J.W., May,G.D. and Dixon, ,R.A.						
REFERENCE	AUTHORS	Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula stem library Unpublished (2000) Contact: Dixon RA Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73402, USA Tel: 580 221 7302					
TITLE	JOURNAL COMMENT						

	Query Match	21.5%	Score 37.2;	DB 94;	Length 671;	
	Best Local Similarity	59.4%	Pred. No. 1,1;			
	Matches	63;	Conservative	0;	Mismatches	43;
					Indels	0;
					Gaps	0
OY	48	ttgcttcctccgcgtcagatctgcttctgctgctcgctagaacctactctgtgc	107			
Db	272	TTCCTTCCGATCGGTCCGTTCCGGTGCTTCATTCTCTTAATCTTACTAATTTGT	213			
OY	108	tggcaggctgcgtctgcttctctctctcaagtctgatc	153			
Db	212	TTCACTTGTTCTTCATTCTCTGCTTACTCTTTCTCGTAAT	167			
RESULT	3					
LOCUS	BF036458/c	733 bp	mRNA	EST	10-OCT-2000	
DEFINITION	601460209p1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3863690 5'					
ACCESSION	BF036458					
VERSION	BF036458.1	GI:10744499				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.					
TITLE	Nih-MGC http://www.ncbi.nlm.nih.gov/MGC/					
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)					
COMMENT	Unpublished (1999)					
	Contact: Robert Strausberg, Ph.D.					
	Tel: (301) 496-1550					
	Email: Robert.L.Strausberg@nih.gov					
	Tissue Procurement: DCTD/DTP					
	cDNA Library Preparation: Life Technologies, Inc.					
	CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)					
	DNA Sequencing by: Incyte Genomics, Inc.					
	Clone distribution: MGC clone distribution information can be					
	found through the I.M.A.G.E. Consortium/LLNL at:					
	http://image.llnl.gov					
	Plate: L10CM591 row: a column: 03					
	High quality sequence stop: 685.					
FEATURES	Location/Qualifiers					
source	1..733					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/clone="IMAGE:3863690"					
	/clone_lib="NIH_MGC_66"					
	/tissue_type="adenocarcinoma"					
	/lab_host="DH10B (phage-resistant)"					
	/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;					
	Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.					
	Average insert size 1.8 kb. Library constructed by Life					
	Technologies."					
BASE COUNT	175 a 195 c 155 t					

ORGANISM	Hordeum vulgare Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.
REFERENCE	Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum. 1 (bases 1 to 743)
AUTHORS	Wing,R., Close,T.J., Kleinholz,A., Wise,R., Begum,D., Frisch,D., Yu, .Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo .T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and Wood,T..
TITLE	Development of a genetically and physically anchored EST resource for barley genomics
JOURNAL	Unpublished (2000)
COMMENT	Contact: Wing RA Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel.: 864 656 7288 Fax: 864 656 4293 Email: rwing@clemson.edu Seq primer: AATTACCCCTCACTAAAGCG High quality sequence start: 148 High quality sequence stop: 584. Location/Qualifiers 1..743 /organism="Hordeum vulgare" /cultivar="ClI6151 (M1ab)" /db_xref="taxon:4513" /clone="HV_CED0016H02F" /library="Hordeum vulgare seedling green leaf EST library HVCNDA0005 (Erysiphe infected & control) " /tissue_type="seedling green leaf" /lab_host="SOLR" /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT	24 a 256 c 88 g 363 t 12 others
ORIGIN	
Query Match	20.2%; Score 35; DB 109; Length 743;
Best Local Similarity	55.3%; Pred.No. 4.7;
Matches	68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
Oy	ctcctgtcgcttcgccttccctcccagatcgctgcttgcttgttcgtacgaac 95
Dd	392 CTGTGTCCTTTGGCGTCTCTCTCTCTTGTAACCCTCGATGTGTCTCTGAATGC 451
Oy	cctaactcgtgatggaggtagtcgcatctgcgttcctaccctccaagtcaagtcagatg 155
Dd	452 COTCCTTTGTCGCCGCTTCTCTTGTGGCGTCTCTGTTGTTGTTGGTTTCTTT 511
Oy	156 gtg 158 Dd 512 TTG 514
RESULT 7	
LOCUS	AM974965 632 bp mRNA EST 02-JUN-2000
DEFINITION	Esf38r070 MAGGE resequences, MAGN Homo sapiens CDNA, mRNA sequence.
ACCESSION	AM974965
VERSION	AM974965.1 GI:8166168
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 632) Begde,P., Ol.R., Abernathy,K., Dharap,S., Gaspar,d.R., Gay,C., Holt .J.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and Quackenbush,J.
TITLE	Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cdna microarray
JOURNAL	Unpublished (2000)
COMMENT	Contact: John Quackenbush The Institute for Genomic Research

						9712 Medical Center Dr., Rockville, MD 20850, USA	
						Tel.: 301 838 3528	
						Fax: 301 838 0208	
						Email: johngetlgr.org	
						Plate: 343	
						Seq primer: Forward.	
FEATURES						Location/Qualifiers	
SOURCE						1..632	
						/organism="Homo sapiens"	
						/db_xref="taxon:9606"	
						/clone_id="MAGE_resequences_MAGN"	
						/note="Vector: pbluescriptSkm"	
BASE COUNT	88 a	163 c	105 g	276 t			
ORIGIN							
Query Match		19.9%	Score 34.4;	DB 97;	Length 632;		
Best Local Similarity		52.9%;	Pred. No. 6.9;				
Matches 74;	Conservative	0;	Mismatches	66;	Indels	0;	Gaps 0;
OY 34	agctcttgctgcgtgcttcttcctcgcacagatcgcttgcttgctgcgtaga	93					
Db 447	ATCTCCCTCCTCGTCGTTCTTCTTCCTTGACGTCAAGACTGTTGGTCCCCTCC	506					
OY 94	accctacctgctgctgcgaagtgcgtctcttcgtcttccctccaattgatcatgt	153					
Db 507	TTCGTGTTCCCTCCCGAAGTGCTCTTGGCCAACCTTCTCTGAGACGTCGTTC	566					
OY 154	gttgtgtgtgggggagcgcag	173					
Db 567	TTCGTGTGATGATGGGGCTG	586					
RESULT 8	CNSO4YVF	990 bp	DNA	GSS	26-JUL-2000		
CNSO4YVF/c	Tetraodon nigroviridis genome survey sequence T7 end of clone						
LOCUS	03E23 of library A from Tetraodon nigroviridis, genomic survey						
DEFINITION	sequence.						
ACCESSION	AL131476	GI:9546360					
VERSION	AL131476.1						
KEYWORDS	GSS: genome survey sequence.						
SOURCE	Tetraodon nigroviridis.						
ORGANISM	Tetraodon nigroviridis						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
	Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;						
	Acanthomorphi; Acanthopterygii; Percormorpha; Tetraodontiformes;						
	Tetraodontidae; Tetraodon.						
REFERENCE	1 (bases 1 to 990)						
AUTHORS	Roeft Croillius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brotilier,P., Quetlier,F., Saurin,W. and Weissbach,J. Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence Nat. Genet. 25 (2), 235-238 (2000) 20296633						
TITLE	2 (bases 1 to 990)						
JOURNAL	Croillius,H.R., Jallion,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetlier,F., Saurin,W., Benoit,A. and Weissbach,J. Characterization and repeat analysis of the compact genome of the freshwater puiterfish tetraodon nigroviridis Genome Res. 10 (7), 939-949 (2000) 20359837						
MEDLINE	3 (bases 1 to 990)						
REFERENCE	Genoscope.						
AUTHORS	Direct Submission						
TITLE	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases						
JOURNAL	This sequence is a single read and was generated as part of a large						
COMMENT	scale clone-end sequencing project of the tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon .						
FEATURES						Location/Qualifiers	

```
source
1. .990
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="053E23"
/clone_lib="A"
/note="Genoscope sequence ID : COA053AC1201-end : 17"

BASE COUNT      292 a      212 c      256 g      210 t      20 others
ORIGIN

Query Match
Best Local Similarity 51.7%; Score 34.2; DB 193; Length 990;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

OY 18 agagtaagcagactagagctctctgctgctgctctctctccgcctcagatctgctgct 77
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 574 AGTAGAGAGAGACCGCGGCTTTCTTCNTCCCTTCTTGTTATATCTCTTACT 515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 78 tgcctgcttcgtagaacctactctgctgctgagtgctgcgctctgctctctccct 137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 TGCCCTCTTTCTCTCCCTCCACACGCTCCCTGGCGTGCCTTTAAATCTTCCTCCCT 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 138 caagtcgactgagtgtgtgtg 162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 454 TCTTCCCGCTTGTTTTGTCTTG 430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
BE268416 702 bp mRNA EST 13-JUL-2000
LOCUS 601124712F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:2989466 5',
DEFINITION mRNA sequence.
ACCESSION BE268416
VERSION BE268416.1 GI:9142021
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 702)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM80 row: g column: 03
High quality sequence start: 8
High quality sequence stop: 173.
Location/Qualifiers
1. .702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2989466"
/clone_lib="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAACG(c). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```
BASE COUNT      328 a      142 c      190 g      42 t
ORIGIN

Query Match
Best Local Similarity 19.5%; Score 33.8; DB 106; Length 702;
Matches 62; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

OY 47 ctgctctctctccgcctcagatctgctgctgctgctgctgtagaacctactctg 106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 396 CGTGCCTTGTCTCTCTCGGCTCGATCTCTGTTTCCCTGCTGCTCCGCT 337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 107 ctgagagtcgctgctgctgctctctctcctcgaatcgatctgt 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 CGTGCCTTCTGATTTGATGTTCTCTCTCGTGACATCTCTTTT 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
BE798838 915 bp mRNA EST 20-SEP-2000
LOCUS 601384668F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3939112 5',
DEFINITION mRNA sequence.
ACCESSION BE798838
VERSION BE798838.1 GI:10220036
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 915)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM787 row: g column: 17
High quality sequence start: 9
High quality sequence stop: 795.
Location/Qualifiers
1. .915
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3939112"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAACG(c). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```
BASE COUNT      282 a      200 c      276 g      157 t
ORIGIN

Query Match
Best Local Similarity 53.4%; Score 33.8; DB 135; Length 915;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

OY 39 ctgcgcgctgctctctccgcctcagatctgctgctgctgctgtagaacct 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 798 CTTTCTCTTCTGCTTTTCATGTTTCGCTACAGATTTCGGGTGCTTCATCTGCATC 739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 99 actctgctcgcagatgctgctgctctctctctcctcaagtcgatctgltg 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


JOURNAL
COMMENT

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/DTP

Tissue Procurement: DCTD/DTPF

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arranged by: The I M A G E Consortium

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM579 row: d column: 16
High quality fluorescence scan: 133

High quality sequence stop: 133.

FEATURES

source

1. .1265

1. .1265

```
/organism="Homo sapiens"
/ab_xref="+axon.9606"
```

```
/db_xref="taxon:9606"
/cclone="TMAGF.3859167"
```

```
/clone="IMAGE:385916/"
```

```
/clone_id="NIH_MGC_66"
/tissue_type="adrenocarc
```

```
celltype="adenocarcinoma"
lab host="DH10B (phage-resis
```

```
/lab_host= drive (phage-resistant)
/notes="Organ: ovary; Vector: PCMV-S
```

site 2: salI: cloned unidirectionally. primer: oligo dT.

Average insert size 1.8 kb. Library constructed by Life

Technologies.

BASE COUNT	662 a	248 c	265 g	89 t	1 others
ORIGIN					

ORIGIN

Query Match

19.38; Score 33.4; DB 139; Length 1265;
53.48; Pred. No. 15;

Matches 70; Conservative 0; Mismatches

Matches 70; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

13 **tgcttagctaacgcactagagcttctgtcgcttgctctctccgcctcagatctgc** 72

[illegible]

QY 73 ttgctctgcttgcgtacgaaccctacatgtgtcgtcgagtgctgcctgctctcttc 132

[illegible]

QY 133 ttcctcaagtt 143

Db 984 TTCGTCCTGTT 974

```
Search completed: February 12, 2001, 21:02:19
Job time: 3975 sec
```

Job time: 3975 sec

THIS PAGE BLANK (USPTO)